بسم الله الرحمن الرحيم

دانشکده مهندسی کامپیوتر

پایان نامه کارشناسی ارشد مهندسی هوش مصنوعی

اندازه‌گیری انسجام یک مکالمه با استفاده از تکنیک‌های هوش مصنوعی

نگارنده:

فاطمه زکی‌زاده قریه‌علی

استاد راهنما:

دکتر مرتضی زاهدی

تیر ۱۳۹۷

شماره: ۵۷۳ ر ف ک
تاریخ: ۹۷/۴/۲۵

باسمه‌تعالی

دانشگاه صنعتی شاهرود

مدیریت تحصیلات تکمیلی

فرم شماره (۳) صورتجلسه نهایی دفاع از پایان نامه دوره کارشناسی ارشد

با نام و یاد خداوند متعال، ارزیابی جلسه دفاع از پایان نامه کارشناسی ارشد خانم **فاطمه زکی‌زاده قریه‌علی** با شماره دانشجویی **۹۴۰۸۸۶۴** رشته مهندسی کامپیوتر گرایش هوش مصنوعی و رباتیک تحت عنوان **اندازه-گیری انسجام یک مکالمه با استفاده از تکنیک‌های هوش مصنوعی** که در تاریخ **۹۷/۴/۲۵** با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح ذیل اعلام می‌گردد:

قبول (با درجه: بسیار خوب) ☐ **مردود** ☐

نوع تحقیق: **نظری** ■ **عملی** ☐

عضو هیأت داوران	نام ونام خانوادگی	مرتبهٔ علمی	امضاء
۱ـ استاد راهنمای اول	دکتر مرتضی زاهدی	استادیار	
۴- نماینده تحصیلات تکمیلی	دکتر فاطمه جعفرنژاد	استادیار	
۵- استاد ممتحن اول	دکتر مرضیه رحیمی	استادیار	
۶ـ استاد ممتحن دوم	دکتر وحید ابوالقاسمی	استادیار	

نام و نام خانوادگی رئیس دانشکده: دکترعلی‌رضا الفی

تاریخ و امضاء و مهر دانشکده: ۹۷/۴/۲۵

تبصره: در صورتی که کسی مردود شود حداکثر یکبار دیگر (در مدت مجاز تحصیل) می تواند از پایان نامه خود دفاع نماید (دفاع مجدد نباید زودتر از ۴ ماه برگزار شود).

تقدیم به

پدر، مادر و همسر مهربانم

تقدیر و تشکر:

در این بخش لازم است از زحمات بی دریغ استاد راهنما، دکتر مرتضی زاهدی و همچنین دکتر محمد مهدی حسینی، مهندس مولود آیت، مهندس غزاله مقدم نژاد و دانشجویان آزمایشگاه وب کاوی و شناسایی الگو برای مساعدت ها و همراهی هایشان قدردانی نمایم.

تعهد نامه

اینجانب **فاطمه زکی‌زاده قریه‌علی** دانشجوی دوره کارشناسی ارشد رشته هوش مصنوعی دانشکده کامپیوتر و فناوری اطلاعات دانشگاه صنعتی شاهرود نویسنده پایان‌نامه **اندازه‌گیری انسجام یک مکالمه با استفاده از تکنیک‌های هوش مصنوعی** تحت راهنمائی **آقای دکتر مرتضی زاهدی** متعهد می‌شوم.

- تحقیقات در این پایان‌نامه توسط اینجانب انجام‌شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورداستفاده استناد شده است.
- مطالب مندرج در پایان‌نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود است و مقالات مستخرج با نام «دانشگاه صنعتی شاهرود» و یا «**Shahrood University of Technology**» به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تأثیرگذار بوده‌اند در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافتهای آن‌ها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

امضای دانشجو **تاریخ**

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود است. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان‌نامه بدون ذکر مرجع مجاز نمی‌باشد.

چکیده

مکالمه یک نوع ارتباط دو طرفه بین دو انسان است. از نظر متخصصین روان‌شناختی مکالمه‌ی صحیح دارای ویژگی‌هایی است که یکی از آن‌ها انسجام می‌باشد. انسجام در مکالمه به معنای حفظ مسیر، عدم شکاف و تغییرات ناگهانی موضوع مکالمه است. تغییر موضوع در یک مکالمه‌ی منسجم به‌صورت روان و مرتبط با موضوع‌های گذشته اتفاق می‌افتد.

یک مکالمه می‌تواند بین انسان و سیستم برقرار شود، مانند سیستم‌های گفتگو و پرسش و پاسخ تعاملی. به منظور افزایش کارایی و محبوبیت سیستم‌های گفتگو، طراحی این سیستم‌ها باید به‌گونه‌ای باشد که معیارهای مکالمه‌ی صحیح بین دو انسان، در آن‌ها رعایت شود. درواقع سیستم‌های گفتگو باید مانند یک انسان خبره پاسخ دهند. مکالمه‌ی صحیح بین دو انسان دارای ویژگی‌هایی است که رعایت آن‌ها در سیستم‌های گفتگو، می‌تواند انسجام را افزایش دهد. ویژگی‌ها از طریق مقایسه‌ی مکالمه‌های منسجم و غیر منسجم قابل شناسایی هستند. با شناخت ویژگی در مکالمه‌های منسجم و غیر منسجم، می‌توان میزان انسجام یک مکالمه را اندازه‌گیری کرد.

تاکنون در سیستم‌های مکالمه‌ای ساخته شده، ذکر شده انسجام در آن‌ها رعایت شده است. طراحان این سیستم‌ها از نظرات انسان برای بررسی میزان رعایت انسجام در مکالمه‌ها استفاده کرده‌اند. استفاده از نظر انسان مفید است، اما انسان نمی‌تواند با یک نظر یکسان به سیستم‌ها امتیاز بدهد. تاکنون سیستم خودکاری طراحی نشده است که بتواند با یک نظر یکسان به مکالمه‌ها امتیاز بدهد.

در این پژوهش با بررسی مکالمه‌های منسجم و غیر منسجم تولید شده توسط انسان، ویژگی‌هایی برای ارزیابی میزان انسجام مکالمه پیشنهاد شده است. ویژگی‌ها به دو دسته تکرار و همنشینی کلمه-های کلیدی تقسیم شده‌اند. هر دسته از ویژگی‌های معرفی شده روی مجموعه مکالمه‌ها مورد آزمون قرار گرفته‌اند. معیار اندازه‌گیری f حاصل از دسته‌بندی مکالمه‌ها با ویژگی‌های مربوط به تکرار کلمه-های کلیدی با ماشین بردار پشتیبان ٪۶۲/۸ می‌باشد. ترکیب تکرار و همنشینی کلمه‌های کلیدی معیار اندازه‌گیری f را به ٪۶۴ رسانده است.

کلمات کلیدی: انسجام در مکالمه، تکرار کلمه‌های کلیدی در مکالمه‌ی منسجم، تأثیر همنشینی کلمه‌ها در انسجام، تولید مکالمه‌ی بدون انسجام

فهرست مطالب

عناوین **شماره صفحه**

فصل اول: ۱
مقدمهای بر انسجام در مکالمه ۱
۱-۱- پردازش زبان طبیعی ۲
۱-۲- انسجام ۲
۱-۳- انسجام در متن ۳
۱-۳-۱- انسجام در خلاصه سازی ۴
۱-۳-۲- انسجام در ساده سازی متن ۴
۱-۳-۳- انسجام در ترجمه آماری ماشینی ۵
۱-۴- انسجام در مکالمه ۵
۱-۴-۱- انسجام در سیستم های گفت‌وگوی وظیفه گرا ۷
۱-۴-۲- انسجام در ربات های مکالمه گرا ۸
۱-۵- ضرورت اندازه‌گیری خودکار انسجام در مکالمه ۸
۱-۶- هدف پایاننامه ۹
۱-۷- فرضیات مسئله ۹
۱-۸- ساختار پایاننامه ۱۰
فصل دوم: ۱۱
مروری بر کارهای گذشته ۱۱
۲-۱- مقدمه ۱۲
۲-۲- انسجام بر پایه موجودیت ۱۲
۲-۳- انسجام لغوی ۱۴
۲-۳-۱- اندازه‌گیری انسجام با LSA ۱۵
۲-۳-۲- اندازه‌گیری انسجام لغوی به وسیله گراف ۱۵
۲-۳-۳- اندازه‌گیری انسجام لغوی با شبکه عصبی ۱۷
۲-۳-۴- دیگر روشهای انسجام لغوی ۱۸
۲-۴- روشهای دیگر اندازه‌گیری انسجام ۲۰
۲-۵- جمع‌بندی ۲۲
فصل سوم: ۲۳

پایگاه داده و استخراج ویژگی ۲۳
(روش پیشنهادی) ۲۳
۳-۱- مقدمه ۲۴
۳-۲- پایگاه داده ۲۴
۳-۲-۱- میزان انسجام مکالمه ۲۵
۳-۲-۲- ساخت پایگاه داده نهایی ۲۸
۳-۳- ویژگیهای پیشنهادی ۲۹
۳-۳-۱- خوشه بندی مکالمه ۳۰
۳-۳-۲- معیارهای استفاده شده در خوشهبندی ۳۲
۳-۳-۳- میزان تکرار کلمه ها در مکالمه ۳۳
۳-۳-۴- بیشترین تعداد تکرار کلمه ها ۳۴
۳-۳-۵ تکرار کلمه با توجه به طول مکالمه ۳۴
۳-۳-۶- تکرار کلمه در دو نوبت متوالی ۳۵
۳-۴- استفاده از همنشینی کلمه ها ۳۶
۳-۴-۱- خوشه بندی ۳۷
۳-۴-۲- تعداد کلمه های همنشین با تاریخچه مکالمه ۳۷
۳-۴-۳- میانگین همنشینی ۳۸
۳-۴-۴- شناسایی شکست های مکالمه ۳۹
۳-۴- دستهبندی مکالمه ها ۴۰
۳-۴-۱- بیزین ساده ۴۱
۳-۴-۲-ماشین بردار پشتیبان ۴۱
۳-۴-۳- درخت تصمیم ۴۲
۳-۵- جمعبندی ۴۲
فصل چهارم: ۴۳
نتایج و آزمایشها ۴۳
۴-۱- مقدمه ۴۴
۴-۲- انتخاب ویژگی ۴۴
۴-۲-۱ مقداردهی مقادیر ثابت ۴۵
۴-۲-۲- انتخاب بهترین ویژگی ها ۴۷
۴-۳- ویژگیهای ارتباط معنایی ۴۹

۱-۳-۴- تعیین مقادیر ثابت ۵۰
۴-۳- همبستگی اندازه‌ی انسجام و ویژگی ها ۵۴
۴-۳- دستهبندی مکالمه ۵۷
۴-۳-۱- نتایج حاصل از دسته بندی ۵۸
۴-۵- جمع‌بندی ۶۲
فصل پنجم: ۶۵
جمع‌بندی و پیشنهادهایی برای ادامه کار ۶۵
۶-۱- جمع‌بندی ۶۶
۶-۲- کارهای آینده ۶۷
منابع و مأخذ: ۶۸

فهرست شکل‌ها

عناوین **شماره صفحه**

شکل (۱-۱): شکل های مختلف انسجام ۳
شکل (۲-۱) نمونه‌ای نمایش موجودیت دو جمله[۲۰] ۱۳
الف) مکالمه با انسجام متوسط (تغییر در نوبت های ۵ و ۹) ۲۷
شکل(۳-۱): نمونه ای از مکالمه دارای انسجام متوسط ۲۷
شکل (۳-۲): نمونه ای از مکالمه بدون انسجام ۲۸
شکل (۴-۱): مکالمه با انسجام عالی ۶۱
شکل (۴-۲): نمونه مکالمه‌ی شماره ۲۰۱۹ ..... ۶۲

فهرست جدول‌ها

عناوین شماره صفحه

جدول (۴-۱): بهره اطلاعات ویژگی های شماره ۳ و ۴ با حداقل β ۴۶
جدول (۴-۳): بهره اطلاعات میانگین تکرار کلمه ها باارزش β نسبت به عمق تکرار ۴۶
جدول (۴-۴): بهره اطلاعات آنتروپی خوشه ها در خوشه بندی با معیار کسینوسی ۴۷
جدول (۴-۵): بهره اطلاعات میانگین خوشه ها در خوشه بندی با معیار کسینوسی ۴۷
جدول (۴-۶): بهره اطلاعات ویژگی شماره ۲۴ با حداقل همنشینی θ ۵۰
جدول (۴-۷): بهره اطلاعات ویژگی های شماره ۲۶ و ۲۷ با حداقل همنشینی θ ۵۱
جدول (۴-۸): بهره اطلاعات ویژگی های شماره ۲۹ و ۳۰ با حداقل همنشینی θ ۵۱
جدول (۴-۹): بهره اطلاعات ویژگی شماره ۳۱ با حداقل همنشینی θ ۵۲
جدول (۴-۱۰): بهره اطلاعات ویژگی شماره ۳۲ با حداقل همنشینی θ ۵۲
جدول (۴-۱۱): بهره اطلاعات ویژگی شماره ۳۳ با حداقل همنشینی θ ۵۲
جدول (۴-۱۳): بهره اطلاعات ویژگی ها ۵۴
جدول (۴-۱۴): میزان همبستگی بین اندازه انسجام و ویژگی ها ۵۷
جدول (۴-۱۵) : نتایج دسته بندی مجموعه های آزمون ۵۹
جدول (۴-۱۶): مقایسه نتایج پیش‌بینی الگوریتم دسته بند روی مجموعه های دوم و سوم ۶۰

فصل اول:

مقدمه‌ای بر انسجام در مکالمه

۱-۱- پردازش زبان طبیعی

وسعت اطلاعات دیجیتال و نیاز به دسترسی سریع اطلاعات، باعث به ایجاد سیستم‌هایی مانند پرسش و پاسخ[۱]، پرسش و پاسخ تعاملی[۲]، خلاصه‌سازی[۳]، سیستم‌های گفتگو[۴] و غیره در حوزه‌ی پردازش زبان طبیعی شده است. پردازش زبان طبیعی[۵] یک روش محاسباتی است که به تحلیل خودکار و نمایش زبان انسان می‌پردازد. پردازش زبان طبیعی، سیستم را قادر می‌سازد تا زبان طبیعی را درک و بتواند گفته‌هایی مشابه با زبان انسان، تولید کند [۱]. استفاده از زبان انسان برای تعامل، این امکان را برای کاربران فراهم می‌کند که به‌آسانی بتوانند با سیستم ارتباط برقرار و اطلاعات مورد نیاز خود را از طریق این سیستم‌ها جستجو کنند.

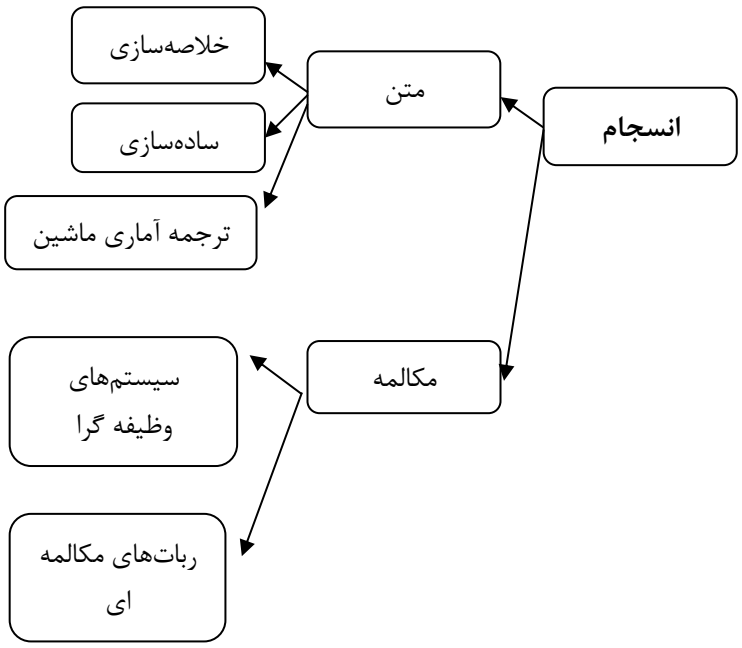
برای کارایی بیشتر و بهتر این سیستم‌ها لازم است، تولیدات آن‌ها مشابه تولیدات انسان خبره در آن حوزه باشد. هرچه شباهت بیشتر باشد، سیستم کارایی و محبوبیت بیشتری خواهد داشت. به‌عبارت‌دیگر، هر چه تعامل طبیعی‌تر باشد رضایت کاربر از سیستم بیشتر است. جمله، پاراگراف و متن‌های تولید شده توسط انسان خبره دارای ویژگی‌هایی است. با شناخت این ویژگی‌ها و اجرای آن‌ها در سیستم‌های مختلف پردازش زبان طبیعی، باعث بهبود کارایی این سیستم‌ها می‌گردد. یکی از این ویژگی‌ها انسجام است. که در ادامه به‌طور کامل، توضیح داده می‌شود.

۱-۲- انسجام

پردازش زبان طبیعی در حوزه‌های ترجمه آماری ماشینی[۶]، سیستم‌های پرسش و پاسخ تعاملی، خلاصه‌سازی، ربات گفتگو[۷] و غیره تعریف شده است. انسجام در هر حوزه تعریف‌ها و معیارهای متفاوتی دارد. رعایت کردن این معیارها در هر حوزه باعث بهبود کارایی سیستم و نزدیک‌تر شدن

[1] Question Answering
[2] Interactive Question Answering System(IQA)
[3] Summarization
[4] Dialog System
[5] Natural Language Processing
[6] Statistical Machine Translation
[7] Chatterbots - Chatbots

خروجی آن به گفته‌های انسان می‌شود. حوزه‌هایی که در پردازش زبان طبیعی کار می‌کنند را می‌توان از نظر انسجام به دو دسته متن و مکالمه طبقه‌بندی کرد، که هر دسته تعریف یکسانی از انسجام دارند. سیستم‌هایی مانند ترجمه آماری ماشینی، خلاصه‌سازی و ساده‌سازی متن، در دسته‌ی متن و سیستم‌های گفتگو، در حوزه مکالمه دسته‌بندی می‌شوند(شکل (۱-۱)).



شکل (۱-۱): شکل‌های مختلف انسجام

۱-۳- انسجام در متن

متنی منسجم است که بین بخش‌های مختلف آن ارتباط منطقی وجود داشته باشد و خواننده بتواند متن را بهتر درک کند [۲]. هر نوع چیدمانی از جمله‌های پشت سرهم، یک متن منسجم را نمی-سازند. جمله‌ها بایستی موضوع یکسان داشته و به یکدیگر نزدیک باشند[۳]. همچنین نحوه ترکیب جمله‌ها در تولید متن منسجم بسیار تأثیرگذار است. پس جمله‌ها باید به هم مرتبط بوده و با ترتیب خاصی کنار هم قرار گیرند. انسجام در متن را می‌توان به سه دسته کلی انسجام در خلاصه‌سازی، ساده‌سازی متن و ترجمه آماری ماشینی، دسته‌بندی کرد.

۱-۳-۱- انسجام در خلاصه‌سازی

خلاصه‌سازی به معنای ایجاد یک متن کوتاه‌تر از یک یا چند مجموعه متن است که مفهوم آن را نمایش می‌دهد. به‌طورکلی روش‌های خلاصه‌سازی به دو دسته خلاصه‌سازی مبتنی بر چکیده[1] و خلاصه‌سازی مبتنی بر استخراج[2] تقسیم‌بندی می‌شوند[۴]. خلاصه‌سازی مبتنی بر چکیده به معنای ایجاد یک خلاصه دقیق برای کاربر است. این نوع خلاصه‌سازی برای سیستم بسیار پیچیده است، زیرا سیستم باید متون مختلف را درک کند و مفهوم متن را تفسیر کند. خلاصه‌سازی مبتنی بر چکیده به دو دسته ساختاری[3] و معنایی[4] طبقه‌بندی می‌شود. خلاصه‌سازی مبتنی بر استخراج، به معنای انتخاب بخش‌های مهم یک یا چند متن توسط سیستم است که به‌طور مختصر هدف متن‌ها را به خواننده می‌رساند. برای انتقال صحیح محتوای متن به خواننده لازم است در ساخت خلاصه، معیارهای انسجام رعایت شود. انتخاب و چینش متوالی جمله‌ها، یک خلاصه صحیح نمی‌سازد. متن تولید شده باید دارای انسجام باشد، به این معنا که جمله‌های پشت سر هم از نظر مفهوم و گرامر به هم مرتبط باشند[۵, ۶].

۱-۳-۲- انسجام در ساده‌سازی متن

سیستم‌های ساده‌ساز خودکار متن پیچیدگی‌های لغوی و گرامری متن را کاهش می‌دهند تا خواننده متن را راحت‌تر بخواند و درک کند. هنگام ساده‌سازی، حذف و تغییر برخی کلمه‌ها و جمله‌ها از متن باعث از بین رفتن انسجام متن می‌شود. نبود انسجام، درک و فهم متن را برای خواننده سخت می‌کند. هدف این سیستم‌ها کاهش پیچیدگی و روان کردن متن برای خواننده است. به همین دلیل رعایت معیارهای انسجام در این سیستم‌ها مهم می‌باشد. مانند مقاله‌ی [۷] که متن‌های پزشکی را با حفظ انسجام برای افراد غیرمتخصص، ساده و قابل فهم می‌کند.

---

[1] Abstract-based Summerization
[2] Extract-based Summerization
[3] Structured-based
[4] Semantic-based

۱-۳-۳- انسجام در ترجمه آماری ماشینی

همان‌طور که گفته شد در هر متن منسجم جمله‌ها از نظر لغوی و نحوی به هم مرتبط هستند. ایجاد این ارتباط در ترجمه‌های آماری ماشین نیز یک امر مهم می‌باشد[۸]. الگوریتم‌های ترجمه ماشینی، متن را جمله به جمله ترجمه می‌کنند. پس لازم است برای ترجمه بهتر و صحیح‌تر، معیارهای انسجام رعایت شود. به‌عبارت‌دیگر ترجمه به‌گونه‌ای باشد که انسجام متن اولیه در متن ترجمه شده حفظ شود. مقاله‌ی [۹] ایده‌ای ارائه داده است که بر اساس انسجام متن اولیه و موضوع جمله‌های آن ترجمه به‌گونه‌ای انجام شود که انسجام حفظ گردد. در فصل بعد این ایده به‌طور کامل توضیح داده می‌شود.

۱-۴- انسجام در مکالمه

عامل‌های مکالمه‌ای[1] یا سیستم‌های گفتگو سیستم‌هایی هستند که از طریق زبان طبیعی (گفتار، نوشتار و یا هردو) با انسان ارتباط برقرار می‌کنند. تعامل با این سیستم‌ها به‌گونه‌ای است که مشابه مکالمه با یک انسان می‌باشد. نسخه‌ی اولیه‌ی سیستم‌های گفتگو، ربات‌های گفتگو بودند [۱۰] که هدف آن‌ها نسبتاً ساده بود. نقش اولیه این سیستم‌ها این بود که کاربران را با تفکر اینکه در حال ارتباط با انسان هستند، مشابه آزمون تورینگ فریب دهند. به‌مرور، علاقه به سیستم‌های گفتگو از تحقیقات علمی خالص تا کاربرد تجاری این تکنولوژی تغییر پیدا کرد. سیستم‌های گفتگو می‌توانند مختص انجام یک کار خاص (وظیفه‌گرا[2]) و یا صرفا برای سرگرمی (ربات‌های مکالمه‌ای[3]) باشند [۱۱]. مکالمه انسانی دارای ویژگی‌هایی است که شناخت این ویژگی‌ها و رعایت آن‌ها در سیستم‌های مکالمه‌ای باعث افزایش کارایی و بهبود سیستم می‌شود.

[1] Conversation Agent
[2] Task Oriented
[3] Conversation Robot

چهار ویژگی نوبت گرفتن، زمینه، ساختار و رهبری مکالمه در مکالمه‌های انسانی وجود دارند که در هنگام ایجاد یک سیستم گفتگو باید مورد توجه قرار گیرند. همه این خصوصیات گفتگوها به عنوان تجزیه و تحلیل مکالمه[1] در [۱۲] بررسی شده است.

نوبت‌گیری به معنای حفظ ساختار توالی مکالمه است. به این صورت که یک نفر صحبت را آغاز و پس از آن، شخص دیگری نوبت خود را شروع می‌کند.

زمینه‌ی مکالمه، دومین مشخصه‌ای است که باید در هنگام ایجاد سیستم‌های گفتگو مورد توجه قرار گیرد. زمینه‌ی مکالمه معنای موضوع مورد بحث است. به منظور برقراری ارتباط موفق، افراد باید اطلاعاتی داشته باشند که به عنوان پس‌زمینه برای فهم سخنان کسی که نوبت را در دست دارد، استفاده کنند. با شروع از یک زمینه مشترک، شرکت‌کنندگان با اضافه کردن مقدار کمی اطلاعات، یک گفتگوی منسجم را می‌سازند. این مهم است که اطمینان حاصل شود که آن‌ها یکدیگر را درک می‌کنند و مکالمه برای آن‌ها قابل فهم است. گاهی اوقات با استفاده از عباراتی مانند «درسته»، «آهان» یا از طریق سؤالات اضافی، توسط مخاطب اعلام می‌شود که گفتگو را درک می‌کنند. همه این‌ها تضمین می‌کند که مکالمه معنی‌دار است.

ساختار مکالمه یکی دیگر از خصوصیات مکالمه است، شامل اطلاعاتی برای برقراری ارتباط، تعداد و نقش هر مخاطب است. رهبری مکالمه یکی دیگر از ویژگی‌های مهم گفتگو است، که معنی مفهوم ابتکار عمل می‌باشد. رهبری مکالمه نشان می‌دهد چه کسی مکالمه را کنترل می‌کند.

رعایت چهار ویژگی ذکر شده در تولید یک سیستم گفتگو ضروری است. همان‌طور که گفته شد یکی از ویژگی‌ها، زمینه گفتگو می‌باشد که در تولید یک مکالمه منسجم مؤثر است. از نظر روانشناسان هم یکی از معیارهای مکالمه صحیح بین دو انسان انسجام است[۱۳]. انسجام مکالمه را می‌توان به دو دسته‌ی انسجام در سیستم‌های وظیفه‌گرا و ربات‌های مکالمه‌ای دسته‌بندی کرد. انسجام در هر دسته از مکالمه معنای خاصی دارد اما به‌طورکلی انسجام در مکالمه به معنای حفظ مسیر مکالمه و عدم

[1] Conversational Analysis

شکاف در آن است. وجود شکاف و تغییر موضوع ناگهانی مکالمه جذابیت و درک مکالمه را برای طرف مقابل کم می‌کند.

۱-۴-۱- انسجام در سیستم‌های گفت‌وگوی وظیفه‌گرا

هدف سیستم‌های وظیفه گرا مانند Siri[1]، Cortana[2] و Alexa[3] کمک به کاربر در یک کار خاص می-باشد. بعضی از سیستم‌های گفتگوی وظیفه‌گرا که توسط برخی از محققین به عنوان مأموران مکالمه، توسط شرکت‌هایی در زمینه‌هایی مانند خدمات مشتری، سرویس‌های خدماتی، هدایت وب‌سایت، هدایت فروش و پشتیبانی فنی استفاده می‌شوند [۱۴]. علت به‌کارگیری این سیستم‌ها ارائه یک روش طبیعی برای تعامل با سیستم است. به‌عبارت‌دیگر، کاربران سیستم گفتگو معمولا نیاز به آموزش ویژه ای ندارند، زیرا سیستم‌ها برای استفاده، آسان و شهودی هستند. همچنین سیستم DoNotPay که یک ربات حقوق‌دان است، می‌تواند در این دسته قرار گیرد [۱۵].

سیستم‌های پرسش و پاسخ تعاملی را می‌توان در حوزه سیستم‌های وظیفه‌گرا دانست. پروژه QALL-ME نمونه اولیه یک سیستم پرسش و پاسخ تعاملی است که در حوزه گردشگری توسعه یافته است [۱۶]. تمام سیستم‌های وظیفه‌گرا با انجام مکالمه‌ای کوتاه با کاربر، اطلاعات مورد نیاز را از او گرفته و پاسخ مناسب را به کاربر می‌دهند [۱۱]. برای رسیدن به پاسخ مناسب یک و یا چندین ارتباط بین سیستم و انسان می‌تواند برقرار شود.

در این سیستم‌ها اندازه‌ی ارتباط کوتاه‌تر و جواب دقیق‌تر باشد کارایی سیستم بیشتر می‌شود. در طراحی این سیستم‌ها سعی شده با رعایت انسجام، مکالمه کوتاه‌تر و جواب دقیق‌تری تولید شود. سؤال‌های مطرح شده و یا پاسخ‌های تولید شده توسط سیستم، هرچه به محتوای مورد پرسش نزدیک‌تر باشد انسجام بیشتر است، در نتیجه سیستم سریع‌تر به پاسخ مناسب می‌رسد [۱۷]. مانند سیستم [۱۸] که یک سیستم پرسش و پاسخ ژاپنی است. در طراحی این سیستم، سعی شده است،

[1] دستیار مجازی ساخته شده توسط شرکت اپل
[2] دستیار مجازی ساخته شده توسط شرکت ماکروسافت
[3] دستیار مجازی ساخته شده توسط شرکت آمازون

جواب‌هایی تولید کند که با دانش پرسش و پاسخ‌های گذشته بیشترین انسجام را داشته باشد. سیستم‌های پرسش و پاسخ با استفاده از صفحه‌های اینترنتی به سؤال‌های پرسیده شده توسط کاربر پاسخ می‌دهند. در واقع [۱۸] با سیستم پرسش و پاسخ که انسجام در آن رعایت شده باشد، سیستم پرسش و پاسخ تعاملی ساخته است. بیشترین انسجام به معنای بیشترین ارتباط بین جواب تولید شده و پرسش و پاسخ‌های گذشته می‌باشد.

۱-۴-۲- انسجام در ربات‌های مکالمه‌گرا

ربات‌های مکالمه‌گرا برای مکالمه‌های طولانی طراحی شده‌اند. موضوع مکالمه با این ربات‌ها مشخص نمی‌باشد و در کار خاصی به کاربر کمک نمی‌کنند. تعامل با این سیستم‌ها به‌گونه‌ای است که مشابه مکالمه با یک انسان می‌باشد. این ربات‌ها جنبه سرگرمی دارند. ELIZA و PARRY نمونه اولیه ربات-های مکالمه‌ای می‌باشند [۱۹].

طراحی این نوع ربات‌ها هرچه بیشتر شبیه به مکالمه انسان باشد، محبوبیت و کارایی آن‌ها بیشتر می-شود. سری جفت از جمله‌ها و یا پرسش و پاسخ‌ها یک مکالمه نمی‌سازند، مکالمه لازم است موضوع داشته باشد [۱۳]. موضوع حین مکالمه می‌تواند تغییر کند اما تغییر باید روان و مرتبط با گفتگوهای گذشته باشد. تغییر موضوع نباید باعث ایجاد شکاف در مکالمه گردد. شکاف و تغییر ناگهانی موضوع از جذابیت مکالمه کم، فهم مکالمه را برای کاربر دشوار می‌کند و باعث نارضایتی کاربر می‌شود. رضایت کاربر در ربات‌های مکالمه‌ای که صرفا برای سرگرمی طراحی شده‌اند بسیار مهم می‌باشد.

۱-۵- ضرورت اندازه‌گیری خودکار انسجام در مکالمه

نبود انسجام در سیستم‌های گفتگو از جمله ربات‌های مکالمه‌ای باعث می‌شود این سیستم‌ها محبوبیت خود را از دست بدهند و کار با سیستم‌های وظیفه‌گرا دشوار شود. چون نمی‌توانند ارتباطی منطقی بین پرسش و پاسخ‌های گذشته پیدا کنند، از مسیر پاسخ اصلی دور می‌شوند. به همین دلیل، در انواع سیستم‌های گفتگوی ساخته شده ذکر شده است که طراحی به‌گونه‌ای بوده که انسجام رعایت شود. نبود سیستمی برای ارزیابی میزان انسجام مکالمه باعث شده برای ارزیابی سیستم از نظرات انسان‌ها

استفاده شود [۱۳]. استفاده از نظر انسان در ارزیابی مفید بوده ولی یک معیار یکسان برای مقایسه با دیگر سیستم‌ها نمی‌باشد. به همین دلیل، ضرورت وجود سیستمی برای ارزیابی میزان انسجام یک مکالمه احساس می‌شود. همچنین می‌توان از سیستم خودکار برای ارزیابی میزان انسجام در زمان ساخت مکالمه استفاده کرد تا پاسخ‌هایی که سیستم می‌دهد منسجم باشد و کارایی سیستم‌های گفتگو افزایش یابد.

۱-۶- هدف پایان‌نامه

در این پژوهش ویژگی‌هایی برای اندازه‌گیری خودکار انسجام مکالمه پیشنهاد داده می‌شود. این ویژگی‌ها بر پایه‌ی تکرار و همنشینی کلمه‌های کلیدی می‌باشند. تاکنون کارهای زیادی برای اندازه-گیری انسجام یک متن انجام شده اما روشی برای اندازه‌گیری انسجام مکالمه معرفی نشده است. در این تحقیق برای اندازه‌گیری انسجام مکالمه از ویژگی‌هایی که برای اندازه‌گیری متن معرفی شده و چند ویژگی پیشنهادی جدید برای اندازه‌گیری انسجام مکالمه استفاده می‌شود. ویژگی‌ها از مکالمه-های مجموعه‌ی آموزش استخراج و در انتها مکالمه‌ها از نظر انسجام به سه دسته‌ی عالی، متوسط و ضعیف دسته‌بندی می‌شوند.

میزان تکرار و ارتباط کلمه‌های کلیدی در مکالمه‌های منسجم با موضوعات متفاوت بسیار نزدیک به هم می‌باشند. ویژگی‌های پیشنهادی و استفاده شده مستقل از مجموعه مکالمه‌های ورودی می‌باشند، زیرا تنها میزان تکرار و همنشینی کلمه‌های کلیدی را در مکالمه‌های منسجم و غیر منسجم بررسی می‌کند.

۱-۷- فرضیات مسئله

در این تحقیق فرض شده جمله موجود در هر نوبت مکالمه انسجام عالی دارد. انسجامی که مورد بررسی قرار می‌گیرد، ارتباط بین گفتگوهای دو طرف است. انسجام محلی که شامل انسجام در یک نوبت مکالمه می‌باشد، بررسی نمی‌شود و فرض شده انسجام آن نوبت عالی است.

۱-۸- ساختار پایان‌نامه

این پایان‌نامه در پنج فصل تدوین شده است. در ادامه‌ی پایان‌نامه در فصل دوم، به معرفی روش‌های اندازه‌گیری انسجام پرداخته می‌شود. فصل سوم روش پیشنهادی معرفی می‌شود. در این فصل نحوه‌ی ساخت پایگاه داده از مکالمه‌ها و انواع ویژگی‌های پیشنهاد شده گفته می‌شود. در فصل چهارم آزمایش‌ها و نتایج بررسی می‌شوند. در نهایت در فصل پنجم، پیشنهادهایی برای بهبود روش ارائه شده و افزایش تشخیص انسجام مکالمه گفته می‌شود.

فصل دوم:

مروری بر کارهای گذشته

۲-۱- مقدمه

همان‌طور که در فصل اول توضیح داده شد انسجام به دو دسته انسجام در مکالمه و انسجام در متن طبقه‌بندی می‌شود. هر دو دسته را می‌توان با تکنیک‌ها و روش‌های یکسانی محاسبه نمود. روش‌هایی که تاکنون برای محاسبه‌ی انسجام استفاده شدند، بر پایه موجودیت، تجزیه تحلیل معنایی نهفته[1]، گراف، شبکه عصبی و غیره هستند. در هر یک از این روش‌ها، تعریف متفاوتی از انسجام مورد بررسی قرارگرفته است. انسجام را برحسب تکرار، همنشینی کلمه‌ها، موضوع جمله‌ها و یا توزیع موجودیت‌ها محاسبه می‌کنند. در این بخش به توضیح مختصری از روش‌های تحلیل انسجام پرداخته می‌شود.

۲-۲- انسجام بر پایه موجودیت

از تحقیقاتی که در تشخیص انسجام انجام شده، استفاده از نوع موجودیت‌های کلمه‌های متن می-باشد. هر کلمه از متن را بر اساس نوع موجودیت آن برچسب می‌زنند. با مقایسه بین توزیع موجودیت-ها بین متن منسجم و غیر منسجم، الگوی متن منسجم را شناسایی می‌کنند.

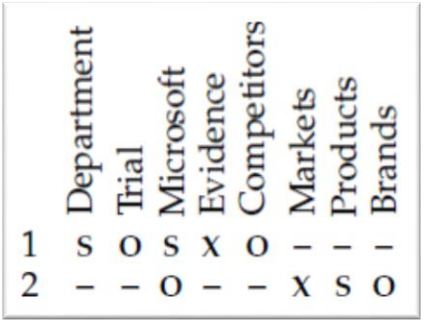
در مقاله‌ی [۲۰] یک مدل انسجام بر پایه موجودیت معرفی شده است. این مدل، الگوهای توزیع موجودیت‌ها را یاد می‌گیرد. در این مقاله معتقد بودند متن منسجم الگوی مرتبی دارد و بین تغییر موجودیت‌ها در جمله‌های مجاور نظم وجود دارد. این مقاله موجودیت‌ها را بر اساس فاعل[2]، مفعول[3]، دیگر دسته‌ها[4] و فاقد موجودیت[5] در هر متن مشخص کرده است. از یک ماتریس دو بعدی برای نمایش توزیع موجودیت‌های متن استفاده کرده‌اند (شکل (۲-۱))

[1] Latent Semantic Analysis (LSA)
[2] Subject
[3] Object
[4] Other
[5] Missing

1- The Justice Department is conducting an anti-trust trial against Microsoft Corp. with evidence that the company is increasingly attempting to crush competitors.

2- Microsoft is accused of trying to forcefully buy into markets where its own products are not competitive enough to unseat established brands.

الف) دو جمله از متن



	Department	Trial	Microsoft	Evidence	Competitors	Markets	Products	Brands
1	S	O	S	X	O	–	–	–
2	–	–	O	–	–	X	S	O

ب)ماتریس موجودیت

شکل (۲-۱) نمونه‌ای نمایش موجودیت دو جمله[۲۰]

همان‌طور که در شکل (۲-۱) نشان داده شده است، ستون ماتریس‌ها، موجودیت‌ها و سطرها، جمله‌ها می‌باشند. با استفاده از نمایش شبکه‌ای نقش نحوی هر موجودیت را ثبت می‌کند و الگوهای توزیع را برای بررسی میزان انسجام به دست می‌آورند. محققان در [۲۱] مدل پیشنهادی برای ارزیابی انسجام را بهبود داده‌اند.

در مقاله‌ی [۲۲] از نمایش مدل توزیع موجودیت‌ها استفاده کرده و با ترکیب ویژگی‌های جدید مرتبط با قوانین گرامری و استفاده از کلمه‌ها به مقالات دانش‌آموزان امتیاز داده است. در مقاله‌ی مذکور نشان داده شده است که استفاده از ویژگی‌های جدید، باعث بهبود تشخیص انسجام می‌شود. معیارهایی که برای بررسی انسجام مورد نظر بوده، خوانایی متن و عدم وجود اطلاعات نویزی و اضافه در متن است. نمایش ماتریسی برای انتقال موجودیت‌ها و به دست آوردن نحوه‌ی توزیع موجودیت‌ها پیچیدگی محاسباتی زیادی دارد. ماتریس، تبدیل به ماتریس اسپارس شده و محاسبات دشوار می‌شود. به همین دلیل در مقاله‌ی [۲۳] برای کاهش پیچیدگی محاسباتی روش ارائه شده را با گراف نمایش می‌دهد. از

گراف برای نمایش و محاسبات انتقال موجودیت‌ها در جملات مجاور استفاده کرده است. استفاده از گراف، بدون تغییر در دقت مدل پیچیدگی محاسباتی را کاهش می‌دهد.

۲-۳- انسجام لغوی

یکی از معیارهایی که برای بررسی انسجام متن یا مکالمه مورد استفاده قرار می‌گیرد انسجام لغوی است. انسجام لغوی یعنی با استفاده از ارتباط بین کلمه‌ها میزان انسجام محاسبه شود [۲۴]. این ارتباط شامل : معنایی، هم‌خانواده، تکرار و غیره می‌باشد. ارتباط بین کلمه‌ها با روش‌های مختلفی به دست می‌آیند. برای به دست آوردن ارتباط معنایی بین کلمه‌ها می‌توان از پایگاه داده‌ای که ارتباط کلمه‌ها را نشان می‌دهند استفاده کرد. پایگاه داده ارتباط معنایی به روش‌های آماری و معنایی (توسط انسان) ساخته می‌شود. در ساخت پایگاه داده معنایی توسط انسان مانند WordNet نوع کلمه از نظر فعل، ضمیر، صفت و همچنین ارتباط کلمه‌ها مشخص شده است. استفاده از روش‌های آماری برای به دست آوردن ارتباط معنایی کلمه‌ها، از روی تعداد متن‌های بسیار زیاد میزان همنشینی و هم‌معنی بودن کلمه‌ها محاسبه می‌شود. مانند روش‌های جانشانی کلمه[1] و تجزیه تحلیل معنایی نهفته، که ارتباط معنایی کلمه‌ها را از روی تعداد متن‌های زیادی به دست می‌آوردند.

ارتباط معنایی و نحوی بین کلمه‌ها برای داشتن متن یا مکالمه منسجم بسیار مهم می‌باشد. زیرا معنی کلمه‌های استفاده شده در متن موضوع را به خواننده می‌رساند. دو کلمه که از نظر معنا به هم مرتبط می‌باشند ممکن است هردو موجودیت آن‌ها یکسان نباشد به همین دلیل انسجام بر پایه موجودیت نمی‌تواند انسجام از نظر معنا را بررسی کند. ارتباط نحوی یعنی کلمه‌ها با ترتیب درستی در جمله‌ها آمده باشند. همچنین میزان تکرار کلمه‌ها و نحوه‌ی توزیع آن‌ها در متن یا مکالمه می‌تواند میزان انسجام را نشان دهد. با مقایسه توزیع و تکرار کلمه‌ها و همچنین میزان ارتباط کلمه‌ها بین متن منسجم و غیر منسجم می‌توان به الگویی از توزیع آن‌ها در متن منسجم رسید. برای اندازه‌گیری انسجام لغوی از روش‌هایی مانند LSA، گراف و شبکه عصبی استفاده شده است.

[1] Word Embedding

۲-۳-۱- اندازه‌گیری انسجام با LSA

LSA یک تکنیک برای به دست آوردن ارتباط معنایی بین کلمه‌ها در چندین هزار متن می‌باشد. LSA هر کلمه منحصر به فرد را با یک بردار چند صد ستونی از رخداد این کلمه در متن‌ها نمایش می‌دهد. به‌طوری که هر دو کلمه که بردار آن‌ها به هم شبیه تر باشد بیشتر باهم در متن یا بخش‌های مختلف آمده‌اند [۲۵]. یک بخش یا کل متن می‌تواند باهمین بردار نمایش داد. هر بردار برای بخش یا متن حاصل میانگین بردار کلمه‌های موجود در متن است. می‌توان با استفاده از نمایش LSA برای بخش‌هایی از متن میزان انسجام متن را محاسبه کرد. در مقاله‌ی [۲۶] بین بردار هر دو بخش (جمله یا پاراگراف) متوالی شباهت کسینوسی آن‌ها را محاسبه و با میانگین گرفتن بین بخش‌ها میزان انسجام متن را محاسبه می‌کند.

در مقاله‌ی [۲۷] از LSA برای محاسبه میزان پیشرفت دانش‌آموزان در آموزش زبان انگلیسی به عنوان زبان دوم آن‌ها استفاده کرده است. محققان در مقاله‌ی مذکور پیشرفت را به معنای انسجام بیشتر در صحبت دانش‌آموزان می‌دانند. نشان داده‌اند با پیشرفت دانش‌آموزان میزان انسجام در صحبت‌های آن‌ها نیز بیشتر می‌شود.

۲-۳-۲- اندازه‌گیری انسجام لغوی به وسیله گراف

در محاسبه انسجام لغوی متن، ارتباط بین جمله‌ها و کلمه‌های متن به روش‌های مختلفی نمایش داده می‌شود. یکی از روش‌های نمایش متن به صورت نمایش گرافی است. در محاسبه میزان انسجام متن هر جمله گره گراف و وزن یال‌ها به عنوان میزان انسجام جمله‌ها در نظر گرفته می‌شود.

در مقاله‌ی [۹] یک روش بر پایه گراف برای خلاصه‌سازی ایمیل ارائه کرده‌اند. این روش با استفاده از سه معیار تکرار کلمه‌های کلیدی[1](رابطه (۲-۱)) ، شباهت معنایی کلمه‌ها بر پایه WordNet (رابطه

[1] Clue words

(۲-۲)) و شباهت دو جمله بر اساس معیار شباهت کسینوسی[1] (رابطه ۲-۳)، میزان ارتباط دو جمله را برای رعایت انسجام در خلاصه‌سازی اندازه‌گیری می‌کند.

$$weight(s_v.s_u) = \sum_{w_i \in s_v} freq(w_i.s_u) \qquad (۱-۲)$$

$$weight(s_v.s_u) = \sum_{w_i \in s_v} \sum_{w_j \in s_u} \sigma(w_i.s_u) \qquad (۲-۲)$$

$$weight(s_v.s_u) = \frac{\overrightarrow{s_u} \cdot \overrightarrow{s_v}}{\|\overrightarrow{s_u}\| \cdot \|\overrightarrow{s_v}\|} \qquad (۳-۲)$$

در رابطه‌ی (۲-۱) وزن بین دو جمله‌ی s_u و s_v ، با استفاده از مجموع تعداد کلمه‌های تکراری (تابع freq) محاسبه می‌شود. رابطه‌ی (۲-۲)، وزن بین دو جمله را با استفاده از مجموع میزان شباهت معنایی کلمه‌های موجود در دو جمله، بر پایه WordNet (تابع σ) محاسبه می‌کند. رابطه‌ی (۲-۳) ابتدا هر دو جمله با یک بردار یکسان که شامل تمام کلمه‌های موجود در دو جمله و تعداد تکرار آن‌ها می‌باشند، نشان داده می‌دهد سپس شباهت کسینوسی بین دو بردار را برای محاسبه‌ی وزن بین دو جمله به دست می‌آورد.

ابتدا گراف جمله‌های مکالمه ساخته می‌شود. هر جمله‌ی مکالمه گره و یال‌ها میزان انسجام دو جمله بر اساس سه معیار ذکر شده به صورت جداگانه می‌باشد. سپس از روی گراف با الگوریتم‌های خلاصه-سازی که قبلا در مقاله‌ی [۲۸] ارائه داده‌اند خلاصه منسجم تولید می‌کنند. در واقع با اضافه کردن انسجام به الگوریتم‌های خلاصه‌سازی، آن‌ها را بهبود می‌بخشند.

در مقاله‌ی [۲۴] از ارتباط معنایی بین کلمه‌ها برای نمایش متن به صورت گراف استفاده شده است. ارتباط معنایی بین کلمه‌ها را از یکی از روش‌های جانشانی کلمه به نام GloVe[2] که در [۲۹] معرفی شده، برای تبدیل هر کلمه به یک بردار چند صد ستونی استفاده کرده است. آن‌ها هر کلمه را به یک بردار ۳۰۰ ستونی نسبت داده‌اند. گراف وزن‌دار را بر اساس ارتباط معنایی بین کلمه‌ها می‌سازند. بیشترین ارتباط معنایی بین کلمه‌های دو جمله، مقدار یال بین دو گره می‌باشد. ارتباط معنایی بین دو جمله را از طریق شباهت کسینوسی بین بردار هر دو کلمه موجود در دو جمله به دست می‌آورد.

[1] Cosine Similarity
[2] Global vectors for word representation

بیشترین شباهت کسینوسی بین دو کلمه در دو جمله، اندازه یال دو جمله می‌شود. محققان در مقاله‌ی [۳۰] یک روش برای به دست آوردن الگویی گرافی که میزان خوانا بودن متن را مدل می‌کند ارائه داده‌اند. برای مدل کردن الگوهای انسجام از روی گراف‌های به دست آمده از متن‌های مختلف، بر اساس زیرگراف‌های هم‌ریخت[1] و الگوی گرافی میزان خوانایی متن، مدل گرافی انسجام را پیشنهاد داده‌اند.

۲-۳-۳- اندازه‌گیری انسجام لغوی با شبکه عصبی

از روش‌ها جدیدی که برای تشخیص انسجام مورد استفاده قرار می‌گیرد می‌توان جانشانی کلمه ، شبکه‌های عصبی و یادگیری عمیق نام برد. جانشانی کلمه روشی جدید برای نمایش یک کلمه در قالب یک بردار می‌باشد. این بردارها می‌توانند به جای کلمه‌ها به عنوان ورودی برای شبکه‌های عصبی مورد استفاده قرار گیرند. در مقاله‌های [۳۱] و [۳۲] که از روش GloVe [۲۹] برای تبدیل جمله‌ها به ورودی شبکه عصبی استفاده کرده‌اند.

در مقاله‌ی [۳۱] با استفاده از یادگیری عمیق مرتب سازی جمله‌ها را به گونه‌ای انجام می‌دهد که متن پایانی منسجم باشد. جمله‌ها را به گروه‌های دو جمله‌ای تقسیم و هر گروه را از نظر انسجام دسته‌بندی می‌کند. در این مقاله از دو روش برای دسته‌بندی استفاده شده است. در یک روش به دو دسته‌ی صفر و ۱ و در روش دیگر به سه دسته ۰،۱ و ۱- تقسیم می‌کند. در روش اول و دوم، ۱ برای حالتی که ترتیب جمله‌ها درست و انسجام داشته باشند و صفر برای حالتی است که ترتیب جملات درست نباشد در نتیجه انسجام ندارد. در روش دوم ۱- برای حالتی است که ترتیب درست باشد اما بین جمله‌ها نیاز به جمله‌ای دیگر باشد تا قابل درک و فهم شوند. سپس هر ترتیب از جمله‌ها که منسجم باشد به عنوان خروجی در نظر گرفته می‌شود. از جمله‌های مقاله‌های موجود در روزنامه‌ها برای پایگاه داده استفاده کرده و حالت‌های غیر منسجم را به صورت خودکار با ترکیب جمله‌ها ساخته است. روش ارائه شده را بر روی دو پایگاه داده آلمانی و انگلیسی آموزش و مورد آزمون قرار داده است.

[1] Isomorphic

در پایگاه داده انگلیسی معیار اندازه‌گیری f برابر ۰/۹۸ و در پایگاه داده آلمانی ۰/۹۷ نتیجه داشته است.

بعضی از روش‌های شبکه عصبی مانند [۳۳] و [۳۴] جانشانی کلمه را در چند لایه اولیه شبکه انجام می‌دهند. این شبکه‌ها با پایگاه‌داده بزرگ آموزش می‌بینند. آن‌ها برای ساخت پایگاه داده غیرمنسجم از ترکیب تصادفی جمله‌ها استفاده می‌کنند. مقاله‌ی [۳۳] سه جمله به عنوان ورودی دریافت می‌کند و با توجه به ترتیبی که دریافت کرده با ۰ یا ۱ در خروجی انسجام را مشخص می‌کند. مقاله‌ی [۳۴] مدل گفتگو با شبکه عصبی تکرار شونده را با افزایش انسجام بهبود می‌دهد. در این روش با یک مکانیزم پویا کلماتی که تولید می‌شوند، بیشترین ارتباط معنایی و نحوی را با تاریخچه گفتگو دارند.

۲-۳-۴- دیگر روش‌های انسجام لغوی

یکی از کاربردهای انسجام در مکالمه می‌توان به دسته‌بندی مکالمه‌های گروهی در مکالمه‌های شبکه-های اجتماعی اشاره کرد. در یک شبکه اجتماعی که تعدادی کاربر در حال مکالمه هستند، ممکن است موضوع در حین مکالمه تغییر کند و یا تعدادی از کاربرها با تعدادی دیگر از کاربرها در حال مکالمه با موضوع متفاوت باشند. روشی برای دسته‌بندی خودکار این نوع مکالمه‌ها بر اساس موضوع پرسش و پاسخ‌های داده شده توسط کاربرها را مقاله‌ی [۳۵] ارائه داده است. آن‌ها با استفاده از شباهت کسینوسی بین جمله‌ها مکالمه‌ها را از نظر موضوع به گونه‌ای دسته‌بندی می‌کنند که در هر دسته انسجام وجود داشته باشد. البته آن‌ها ویژگی‌های دیگری از جمله زمان ارسال پیام، شناسه فرد ارسال کننده پیام و طول پیام هم در دسته‌بندی در نظر داشته‌اند. مسئله‌ای که مورد توجه قرار گرفته است پاسخ‌های متداولی هستند که با هر گروه می‌توانند بیایند. پس پاسخ‌های متداول را شناسایی و در تمام دسته‌ها قرار می‌دهند. برای محاسبه‌ی شباهت بین جمله‌ها، هر جمله را با ضرب مقدار تکرار عبارت (tf) و فرکانس معکوس سند، در یکدیگر نمایش داده است (رابطه(۲-۴)). سپس بین هر دو جمله، شباهت کسینوسی محاسبه می‌شود (رابطه (۲-۵)).

$$SW_{yk} = tf_{yk} * idf_k * coef_k \qquad (۴-۲)$$

$$sim(x.y) = \frac{\sum_{k=1}^{n} SW_{xk} * SW_{yk}}{\sqrt{\sum_{k=1}^{n} SW_{xk}^2 * \sum_{k=1}^{n} SW_{yk}^2}} \qquad (۵-۲)$$

همان‌طور که در رابطه (۲-۴) مشخص است. برای شباهت هر دو جمله از ضرب مقدار تکرار عبارت و فرکانس استفاده کرده است. هنگامی که دو ویژگی تکرار عبارت و فرکانس معکوس سند در یکدیگر ضرب می‌شوند، مجموعه ویژگی جدیدی به وجود می‌آید که قادر است عبارت‌های مفهومی را از عبارت‌های متداول جدا کند. همچنین با مقدار coef شباهت کلمه‌ها را در تعیین میزان شباهت دو جمله به کار برده است. شباهت معنایی فقط در محاسبه یکی از بردارها x و یا y تأثیر داده شود کافی است.

رعایت انسجام در ترجمه‌های ماشینی یک چالش است. همان‌طور که گفته شد الگوریتم‌های ترجمه آماری ماشینی جمله به جمله ترجمه انجام می‌دهند. پس نیاز است بین جمله‌ها ارتباط از نظر موضوع برقرار تا ترجمه یک متن شامل چندین جمله به‌درستی انجام شود. هرچه انسجام متن ترجمه شده بیشتر باشد کارایی سیستم بهتر و به ترجمه توسط انسان خبره نزدیک‌تر است. این انسجام شامل رعایت قواعد گرامری، انسجام موضوعی، انسجام لغوی و نحوی می‌باشد.

یک راه برای حفظ انسجام در زمان ترجمه ماشینی ساخت زنجیره‌ی لغوی متن ترجمه شده از روی زنجیره‌ی لغوی متن اولیه است. زنجیره لغوی از کنار هم قرار گرفتن کلمات به هم مرتبط ساخته می‌شود. زنجیره لغوی یک نمایش از پیوستگی لغوی را فراهم می‌کند. مقاله‌ی [۸] با استفاده از بیشترین آنتروپی زنجیره‌ی لغوی متن اولیه، زنجیره لغوی متن ترجمه شده را می‌سازد.

روشی دیگر برای نمایش ارتباط بین کلمه‌ها استفاده از مقدار PMI[1] بین دو کلمه است [۳۶]. مقدار PMI بین دو کلمه a و b با استفاده از رابطه (۲-۶) به دست می‌آید.

$$PMI(a.b) = log^{\frac{p(a.b)}{p(a)p(b)}} \qquad (۶-۲)$$

[1] Pointwise Mutual Information

P(a) احتمال حضور کلمه‌ی a، P(b) احتمال حضور کلمه‌ی b و P(a,b) احتمال وجود هر دو کلمه در متن یا قسمت‌هایی از متن است که از روی آن‌ها ارتباط کلمه‌ها به دست می‌آید. مقاله‌ی [۳۷]، PMI نرمال شده را پیشنهاد داده است که PMI با تقسیم بر $-log^{p(a.b)}$ مقدار آن نرمال می‌شود (رابطه (۲-۷)).

$$NPMI(a.b) = \frac{log^{\frac{p(a.b)}{p(a)p(b)}}}{-log^{p(a.b)}} \qquad (۲-۷)$$

در مقاله‌ی [۲۵] برای بررسی میزان انسجام متن فقط از مقادیر مثبت NPMI استفاده کرده است. مقدار NPMI هر دو کلمه از متن را باهم جمع کرده و بین آن‌ها میانگین می‌گیرد. در انتها این مقاله، نشان داده است بین مقدارهای به دست آمده و نظر انسان در مورد انسجام متن رابطه مستقیم وجود دارد.

مقاله‌ی [۳۸] معیارهایی بر پایه جانشانی کلمه برای محاسبه انسجام بین موضوع و محتوای توییت ارائه داده است. روش ارائه شده را روی پایگاه داده‌ای عظیم اجرا نموده و بردار کلمه‌ها را به دست آورده است. سپس از بردار کلمه‌ها برای به دست آوردن انسجام بین توییت و موضوع استفاده کرده است.

رابطه (۲-۸) میزان انسجام موضوع توییت را نشان می‌دهد، ارتباط کلمه‌ها (تابع f) را با سه روش LSA، PMI و جانشانی کلمه به دست آورده و نشان داده استفاده از جانشانی کلمه نتیجه‌ی بهتری دارد.

$$\frac{1}{\sum_{m=1}^{n-1} m} \sum_{i=1}^{n} \sum_{j=i+1}^{n} f(w_i . w_j) \qquad (۲-۸)$$

۲-۴- روش‌های دیگر اندازه‌گیری انسجام

همان‌طور که گفته شد رعایت انسجام در ترجمه ماشینی برای ترجمه بهتر و قابل درک یک مسئله، مهم است. هنگامی یک متن حاوی چند جمله به سیستم داده شود مهم است موضوع جمله‌ها شناسایی و بر اساس موضوع آن‌ها ترجمه صورت گیرد. مقاله‌ی [۳۹] برای بهبود انسجام در متن‌های

ترجمه شده توسط ماشین به هر جمله‌ی متن اولیه برچسب موضوع می‌دهد. موضوعات کنار هم یک زنجیره را می‌سازند. بر اساس آنکه متن اولیه متنی منسجم بوده است با زنجیره موضوع آن‌ها زنجیره موضوعی متن ترجمه شده را پیش‌بینی می‌کند. سپس این زنجیره را روی متن ترجمه شده اعمال می‌کند تا متن نهایی هم به‌اندازه متن اولیه منسجم باشد. در واقع انسجام متن اولیه به متن ترجمه شده منتقل می‌شود.

مقاله‌ی [۵] خلاصه‌سازی چند متن را به صورت منسجم انجام می‌دهد و برای استخراج خلاصه و رعایت انسجام از ترکیبی از انسجام لغوی و موجودیت استفاده می‌کند. با یک پایگاه داده از مقاله‌های روزنامه‌ها میزان همنشینی هر دو کلمه را به دست می‌آورد. همچنین با استفاده از WordNet اسم‌ها و فعل‌ها را در متن شناسایی می‌کند و ارتباط بین جمله‌ها را با گراف وزن‌دار نمایش می‌دهد. جمله‌ها، نودهای گراف و وزن یال‌ها، میزان ارتباط بین دو جمله است. معیارهای ارتباط دو جمله که برای ایجاد یال در این مقاله استفاده شده است شامل: می‌توان به موارد زیر اشاره کرد: ۱- در جمله s_1 فعل‌هایی آمده است که مربوط به اسامی است که در جمله‌ی s_2 بوده است. ۲- جمله s_1 به s_2 متصل می‌شود اگر در جمله دوم رخداد یا موجودیت‌هایی باشد که در جمله‌ی اول وجود داشته باشند به‌طوری که زمان رخداد آن بیشتر مساوی جمله دوم باشد. ۳- اگر در جمله‌ها کلماتی مانند اما، اگر، اگر چه و .... باشد، یالی بین دو جمله متوالی قرار می‌گیرد. ۴- در صورتی که s1 با s2 یال داشت و s2 با s3 نیز یال داشته باشد آنگاه از s1 به s3 یال قرار می‌گیرد.

سپس خلاصه متون را از روی گراف ساخته شده، می‌سازد. با استفاده از جمع وزن یال‌ها میزان انسجام متن ساخته شده را بررسی می‌کند.

از روش‌های جدید که برای ایجاد انسجام در تولید خلاصه‌های اتوماتیک پیشنهاد شده می‌توان به [۶] اشاره کرد. این روش با استفاده از یک مدل شناختی، خلاصه‌ای منسجم از یک متن داستانی استخراج می‌کند. در این روش ویژگی‌های افراد، موقعیت و وضعیت شناسایی می‌کند و سپس از روی آن‌ها خلاصه متن را تولید می‌کند.

۲-۵- جمع‌بندی

در این فصل روش‌های معروف و مؤثر در اندازه‌گیری انسجام توضیح داده شد. این روش‌ها شامل LSA، گراف، شبکه عصبی، توزیع موجودیت‌ها و غیره بوده که هر یک کارهایی که روی متن برای اندازه‌گیری انسجام انجام داده‌اند، معیار متفاوتی برای انسجام در نظر گرفته‌اند. سپس با توجه به آن معیار و روش‌های معرفی شده به‌اندازه‌گیری انسجام پرداخته‌اند. معیارهای معرفی شده در اندازه‌گیری انسجام شامل توزیع موجودیت‌ها، چینش، ارتباط معنایی، تکرار و توزیع کلمه‌ها و غیره می‌باشد.

فصل سوم:

پایگاه داده و استخراج ویژگی

(روش پیشنهادی)

۳-۱- مقدمه

در این فصل به چگونگی تولید پایگاه داده و ویژگی‌های پیشنهاد شده برای تجزیه و تحلیل مکالمه‌ها پرداخته شده است. به دلیل عدم وجود پایگاه داده مکالمه‌ای برای ارزیابی انسجام، مکالمه‌های آموزش زبان انگلیسی جمع‌آوری شده‌اند. در این فصل به تولید مکالمه با انسجام‌های مختلف پرداخته شده است.

در تشخیص سطح انسجام یک مکالمه لازم است ویژگی‌هایی از آن مکالمه استخراج شود. با ویژگی‌ها استخراج شده می‌توان به تحلیل مکالمه پرداخت و میزان انسجام را تشخیص داد. در یک مکالمه منسجم چون موضوع ثابت یا تغییرات موضوع روان است، کلمه‌ها تکرار و یا به هم مرتبط می‌باشند. با استفاده از توزیع کلمه‌های تکراری و میزان ارتباط کلمه‌ها می‌توان انسجام یک مکالمه را محاسبه کرد. ویژگی‌هایی که در این پژوهش استفاده شده است، تکرار، نحوه‌ی توزیع و همچنین میزان همنشینی کلمه‌ها در یک مکالمه می‌باشد.

با استفاده از بهترین ویژگی‌های انتخاب شده و الگوریتم‌های دسته‌بندی، یک مکالمه از نظر انسجام به سه دسته خوب، متوسط و ضعیف دسته‌بندی می‌شود.

۳-۲- پایگاه داده

برای ساخت سیستمی که بتواند میزان سطح انسجام یک مکالمه را تشخیص بدهد به پایگاه داده‌ای از چند صد مکالمه با موضوع‌های مختلف احتیاج است. هرچه تنوع موضوع و تعداد مکالمه‌ها بیشتر باشد، سیستم بهتر می‌تواند میزان انسجام را تشخیص دهد. به دلیل عدم وجود پایگاه داده مکالمه‌ای، پایگاه داده‌ای شامل چند صد مکالمه با موضوع‌های مختلف باید ساخته شود. به دلیل آن‌که ساخت مکالمه‌های منسجم زمان‌بر و دشوار است، از مکالمه‌هایی که برای آموزش زبان انگلیسی موجود بوده‌اند استفاده کرده‌ایم. ۲۴۲۰ مکالمه که برای آموزش زبان انگلیسی بوده‌اند از سایت www.easypacelearning.com با موضوع‌های مختلف جمع‌آوری شده است. مکالمه‌ها محاوره‌ای، بین

دو نفر و با موضوع‌های مختلف مانند اجاره کردن خانه، چگونه کار پیدا کردن، مسافرت در تعطیلات آخر هفته و غیره می‌باشند.

۳-۲-۱- میزان انسجام مکالمه

تمام ۲۴۲۰ مکالمه جمع‌آوری شده دارای انسجام سطح خوب (سطح A) می‌باشند. برای آنکه ویژگی-های مؤثر در انسجام را روی مکالمه‌ها مورد آنالیز قرار دهیم، به مکالمه‌های بدون انسجام نیز نیاز است. پایگاه داده شامل ۲۴۲۰ مکالمه را به‌طور تصادفی به ۳ سطح انسجام خوب، متوسط (سطح B) و ضعیف (سطح C) دسته‌بندی کرده‌ایم.

مکالمه‌های دارای انسجام کمتر و یا بدون انسجام را می‌توان توسط انسان ساخت و امتیازدهی کرد. ساخت چند صد مکالمه بدون انسجام و یا کمتر از عالی و امتیازدهی آن‌ها توسط انسان کاری زمان‌بر و مشکل است. همچنین انسان نمی‌تواند به همه مکالمه‌ها با یک نظر یکسان امتیاز بدهد. به همین دلیل مکالمه‌های سطح متوسط و ضعیف به صورت خودکار مانند مقاله‌های [۳۳, ۳۴] با ترکیب تصادفی مکالمه‌ها ساخته شده‌اند. برای این کار تعدادی از نوبت‌های هر مکالمه را تغییر می‌دهیم و مکالمه با سطوح پایین‌تر ساخته می‌شود.

۲۴۲۰ مکالمه به صورت تصادفی به سه دسته ۸۰۰ تایی تقسیم شده‌اند. در هر دسته موضوع‌های مختلفی موجود می‌باشد. دسته‌ی اول شامل ۸۰۰ مکالمه بدون تغییر برای مکالمه با سطح خوب می-باشند. در دسته دوم و سوم چند نوبت از هر مکالمه به صورت تصادفی با چند نوبت از مکالمه‌هایی دیگر جانشین می‌شود. تعداد تغییر نوبت‌ها در هر مکالمه و انتخاب مکالمه‌ها برای جانشین شدن نوبت‌ها تصادفی می‌باشد. نوبت اول هیچ یک از مکالمه‌ها به دلیل نوبت اصلی و شروع مکالمه تغییر نمی‌کند. همچنین هنگام تغییر چند نوبت از مکالمه، از چند مکالمه متفاوت استفاده شده است. در مکالمه‌های سطح متوسط و ضعیف به ترتیب کمتر و بیشتر از ٪۳۰ نوبت‌ها تغییر کرده است.

علاوه بر آنکه هر مکالمه بر اساس انسجام به یکی از سه دسته خوب، متوسط و ضعیف دسته‌بندی شده است، امتیاز عددی انسجام هر کدام نیز محاسبه می‌شود. امتیاز داده شده با توجه به تعداد تغییر

نوبت‌های هر مکالمه می‌باشد. تعداد نوبت‌های تغییر نکرده تقسیم بر تعداد کل نوبت‌های هر مکالمه، به عنوان میزان عددی انسجام در نظر گرفته شده است. در رابطه‌ی (۳-۱) N_c تعداد نوبت‌های تغییر نکرده و $|conv_i|$ تعداد کل نوبت‌های مکالمه‌ی i ام می‌باشد.

$$coherence(i) = \frac{N_c}{|conv_i|} \qquad (۳-۱)$$

در ساخت مکالمه با انسجام کمتر این فرض گذاشته شده است که هر نوبت مکالمه از نظر انسجام صحیح است، زیرا مکالمه‌های بدون انسجام از ترکیب مکالمه‌های سطح عالی ساخته شده و تغییری در ترتیب کلمه‌های موجود در نوبت‌های جانشین شده ایجاد نشده است. پس عدم انسجام در مکالمه به دلیل عدم ترتیب درست جمله‌ها و عدم ارتباط با موضوع مکالمه است. نمونه‌ای از مکالمه‌های دارای انسجام متوسط و ضعیف در شکل (۳-۱) و (۳-۲) آمده است.

personA: There are so many flavors of ice cream to choose from!

personB: Yes, there must be about a hundred. Do you have a favorite?

personA: I love fruit flavored ices the best.

personB: Yes, fresh fruit flavored is the best ever.

*personA: Your fees total $235.13.

personB: I don't like peppermint ice cream.

personA: Yes, I know what you mean.

personB: One time, I tried garlic ice cream.

*personA: I appreciate your help. Maybe I'll be back for more computer stuff.

personB: That doesn't sound like a good choice for today. Let's order!

coherence: B 0.20000000298023224

الف) مکالمه با انسجام متوسط (تغییر در نوبت‌های ۵ و ۹)

personA: There are so many flavors of ice cream to choose from!

personB: Yes, there must be about a hundred. Do you have a favorite?

personA: I love fruit flavored ices the best.

personB: Yes, fresh fruit flavored is the best ever.

personA: What is your least favorite ice cream?

personB: I don't like peppermint ice cream.

personA: Yes, I know what you mean.

personB: One time, I tried garlic ice cream.

personA: That could be interesting.

personB: That doesn't sound like a good choice for today. Let's order!

ب) مکالمه منسجم (الف)

شکل(۳-۱): نمونه‌ای از مکالمه دارای انسجام متوسط

personA: I don't know where the EDD office is.

*personB: A Notice of Unemployment Insurance Award letter will be sent to you.

*personA: I know a good Chinese restaurant.

personB: The office has many wonderful workshops to help you find a job.

*personA: Whatever floats your boat.

*personB: The rent each month is $1050.

*personA: I use it rarely. When I do use it, I put pasta or other soft food into it.

*personB: Here you go.

personA: We're just going door to door.

*personB: I really can't stand going to parties.

Coherence: C 0.800000011920929

الف) مکالمه با انسجام ضعیف (تغییر در نوبت‌های ۲، ۳، ۵، ۶، ۷، ۸ و ۱۰)

personA: I don't know where the EDD office is.

personB: The whole process can be done over the phone or Internet. You don't need to go into the office.

personA: What about their job help information?

personB: The office has many wonderful workshops to help you find a job.

personA: Where can I find my local office?

personB: Employment Development Offices are listed in the phone book.

personA: Have you ever seen the office that serves our local area?

personB: The office is two blocks west of the college on Green Street.

personA: If I show up in person, will I get my money any faster?

personB: You have to wait one week before you can collect, regardless of how you file.

ب) مکالمه منسجم (الف)

شکل (۳-۲): نمونه‌ای از مکالمه بدون انسجام

۳-۲-۲- ساخت پایگاه داده نهایی

مکالمه‌هایی که جمع‌آوری شده‌اند مانند هر داده خام دیگر، حاوی ناخالصی‌های زیادی می‌باشند. این ناخالصی‌ها به صورت غلط‌های املائی، حالت‌های اختصار کلمه، حالت‌های نامتعادل و غیره به جمله‌ها

اضافه می‌شوند. در ابتدا باید پردازش‌های اولیه‌ای بر روی منابع خام انجام داد تا بتوان ناخالصی‌ها و نویزهایی که دقت را کاهش می‌دهند را حذف کرد.

علائم نگارشی: وجود علائم نگارشی مانند نقطه و علامت سؤال، اطلاعاتی در مورد اندازه انسجام به ما نمی‌دهد. بهتر است این علامت‌ها پیش از پردازش حذف شوند.

ایست واژه[1]: در هر متنی کلماتی وجود دارد که به تعداد زیاد در متن تکرار می‌شوند و با حذف آن‌ها معنی متن تغییری نمی‌کند. به این نوع کلمه‌ها، ایست واژه گفته می‌شود. به عنوان مثال در زبان فارسی با، در، از، یا، و، به و در زبان انگلیسی on، of، and و or کلمه‌های اضافی هستند. اکنون برای زبان‌های مختلف، مجموعه کلمه‌های اضافه به صورت آماده موجود است. در این پژوهش برای حذف کلمه‌های اضافی از مجموعه کلمه‌های اضافی موجود در کتابخانه استنفورد استفاده می‌شود.

استفاده از شکل‌های یکسان حروف: در بعضی از زبان‌ها برای یک حرف، حالت‌های مختلفی وجود دارد به عنوان مثال در زبان انگلیسی، هر حرف به دو شکل کوچک و بزرگ نمایش داده می‌شود. با در نظر گرفتن تنها یک شکل از حرف می‌توان دقت را بالا برد و از تشخیص اشتباه جلوگیری کرد.

بعد از آماده‌سازی مکالمه‌ها لازم است مجموعه آموزش و آزمون از آن‌ها جدا شوند. با این کار مکالمه-های یکسان و مشخصی برای جداسازی انواع ویژگی‌ها استفاده می‌شود. به‌طور تصادفی از هر دسته ۷۰۰ مکالمه برای آموزش و ۱۰۰ مکالمه برای آزمون جدا شده است. هیچ داده‌ای در بخش آموزش با بخش آزمون مشترک نیست.

۳-۳- ویژگی‌های پیشنهادی

قسمت مهم آنالیز مکالمه استخراج ویژگی‌های مناسب است. ویژگی‌هایی که از مکالمه استخراج می-شوند می‌توانند آماری، معنایی و غیره باشند. در این پژوهش از ویژگی‌های آماری برای آنالیز استفاده شده است. در یک مکالمه منسجم چون موضوع ثابت یا تغییرات موضوع روان است، کلمه‌ها تکرار و یا

[1] Stop Words

به هم مرتبط می‌باشند. از روی توزیع کلمه‌ها و میزان ارتباط آن‌ها می‌توان انسجام را محاسبه کرد. ویژگی‌های استخراج شده شامل خوشه‌بندی و دسته‌بندی مکالمه، تکرار و میزان ارتباط کلمه‌ها در حالت‌های مختلف مانند میانگین ارتباط کلمه‌ها، میزان تکرار کلمه‌ها در دو نوبت متوالی، تکرار کلمه‌ها با توجه به عمق مکالمه و غیره می‌باشند. تعداد ۳۵ ویژگی مؤثر در انسجام مکالمه پیشنهاد داده شده است. ویژگی‌ها در ادامه توضیح داده می‌شوند.

۳-۳-۱- خوشه‌بندی مکالمه

مقاله‌ی [۳۵] یک مکالمه گروهی که چند نفر در حال صحبت هستند را بر اساس موضوع‌های مورد بحث، دسته‌بندی می‌کند. از این کار الهام گرفته‌ایم و با توجه به تکرار و میانگین اندازه ارتباط کلمه‌ها در جمله‌ها، نوبت‌های یک مکالمه را خوشه‌بندی می‌کنیم. تعداد خوشه‌های تشکیل شده میزان انسجام مکالمه را می‌تواند نشان دهد. هر چه تعداد خوشه‌های کمتری تولید شود مکالمه منسجم‌تر می‌باشد، به دلیل آنکه معیارهای مختلف خوشه‌های مختلفی تشکیل می‌دهند. در این پژوهش از معیارهای مختلفی برای خوشه‌بندی استفاده شده است. نحوه‌ی خوشه‌بندی به این گونه است که نوبت اول مکالمه، به آن دلیل که نوبت اصلی مکالمه می‌باشد خود به تنهایی یک خوشه تشکیل می‌دهد. اگر نوبت دوم با توجه به معیارهایی که در نظر گرفته شده به خوشه‌ای اول نزدیک بود، متعلق به آن خوشه می‌شود در غیر این صورت خود، خوشه‌ای جدید ایجاد می‌کند. نوبت‌های بعدی به ترتیب اگر به میزان کافی از نظر معیار استفاده شده به یکی از خوشه‌ها نزدیک بود با آن خوشه ترکیب می‌شود. اگر به هیچ خوشه‌ای نزدیک نبود، خوشه‌ای جدید تشکیل می‌دهد. این کار برای تمام نوبت‌های مکالمه تکرار می‌شود.

در هنگام خوشه‌بندی اگر یک نوبت به چند خوشه به‌طور یکسان نزدیک بود آن چند خوشه ترکیب می‌شوند. وجود جمله‌ای که به دو خوشه نزدیک باشد نشان می‌دهد که با احتمال زیاد دو خوشه از نظر معنای کلمه‌ها به هم نزدیک بوده ولی کلمه‌های مشابه نداشته‌اند. معیارها برای خوشه‌بندی

به‌گونه‌ای در نظر گرفته می‌شود که نوبت‌های مرتبط و منسجم در یک خوشه قرار گیرند. فرض شده نوبت‌های یک مکالمه در صورتی به هم مرتبط هستند، که کلمه‌های کلیدی در تمام نوبت‌ها تکرار شود. بر این اساس معیارهای خوشه‌بندی برحسب تکرار کلمه‌ها می‌باشند.

تعداد خوشه‌های تشکیل شده با طول مکالمه یا همان تعداد نوبت‌های مکالمه رابطه دارد. هرچه تعداد نوبت‌ها بیشتر باشد، تعداد خوشه‌ها هم می‌تواند بیشتر شود و انسجام کاهش نیابد. و در مکالمه‌های کوتاه وجود تعداد خوشه‌های زیاد باعث کاهش انسجام می‌شود. به‌عبارت‌دیگر تعداد خوشه‌های بیشتر به معنای انسجام کمتر است. در صورتی که در مکالمه‌های طولانی تعداد خوشه‌های بیشتر به معنای انسجام کمتر نمی‌باشد؛ به همین دلیل تنها تعداد خوشه‌ها نمی‌تواند به عنوان ویژگی در نظر گرفته شود.

در این پژوهش از دو ویژگی میانگین و آنتروپی[1] خوشه‌ها برای نمایش نتایج خوشه‌بندی استفاده شده است. میانگین خوشه‌ها مشخص می‌کند در هر خوشه به‌طور میانگین چند نوبت وجود دارد. در واقع، تعداد خوشه‌ها با تقسیم بر طول مکالمه نرمال می‌شود(رابطه (۳-۲)). در رابطه (۳-۲) D، مجموعه دسته‌ها و $conv_c$، مکالمه‌ی شماره c ام است.

$$M = \frac{|D|}{|conv_c|} \qquad (۳-۲)$$

علاوه بر مهم بودن تعداد خوشه‌ها تعداد نوبت‌های موجود در هر خوشه نیز در میزان انسجام مکالمه تأثیر دارد. رابطه (۳-۲) نحوه‌ی پخش شدن تعداد نوبت‌ها را در خوشه‌ها نشان نمی‌دهد. به‌طور مثال اگر ۱۰ نوبت داشته باشیم که به ۲ دسته تقسیم شده باشند، آن مکالمه‌ای که به دو دسته‌ی ۵ نوبته تقسیم شده است، انسجام کمتری نسبت به مکالمه‌ای دارد که به دسته‌ی ۹ و ۱ نوبتی تقسیم شده است. با این حال رابطه (۳-۲) برای هر دو حالت عدد یکسانی دارد. به همین دلیل از آنتروپی دسته‌ها (رابطه (۳-۳)) به عنوان ویژگی دیگر استفاده شده است.

[1] Entropy

$$H = \frac{-\sum_{i=1}^{n} \frac{|D_i|}{|conv_c|} * \log \frac{|D_i|}{|conv_c|}}{\log |conv_c|} \qquad (۳-۳)$$

در رابطه (۳-۳) D_i، خوشه‌ی شماره i ام و $conv_c$، مکالمه شماره c ام است. در این رابطه آنتروپی مکالمه‌ی c محاسبه می‌شود. هرچه میزان آنتروپی دسته‌ها بیشتر، انسجام کمتر است. در ادامه به توضیح انواع معیارهای استفاده شده برای خوشه‌بندی پرداخته می‌شود. در پایان خوشه‌بندی هر معیار، دو عدد میانگین و آنتروپی خوشه‌ها به عنوان نتیجه خوشه‌بندی محاسبه می‌شود. شماره ویژگی‌های ۱، ۲، ۲۰، ۲۱، ۲۲ و ۲۳ میانگین و آنتروپی‌های خوشه‌بندی با معیارهای متفاوت می‌باشند.

۳-۳-۲- معیارهای استفاده شده در خوشه‌بندی

در ویژگی‌های ۱ و ۲ بر اساس تکرار کلمه‌ها خوشه‌بندی انجام می‌شود. در هر خوشه با نوبت‌هایی که در آن خوشه قرار دارند، یک کیسه از کلمه‌ها ساخته می‌شود. هر نوبت مکالمه، متعلق به خوشه‌ای است که بیشترین کلمه‌های یکسان را با کیسه‌ی کلمه‌های آن خوشه داشته باشد. اگر چند خوشه به‌طور یکسان تعداد کلمات مشابه داشته باشند، آن چند خوشه باهم ترکیب می‌شوند. هنگامی که هر نوبت به خوشه‌ای تعلق می‌گیرد، کیسه‌ی کلمه‌ها به روز می‌شود. کلمه‌های نوبت جدید در کیسه‌ی کلمه‌های خوشه‌اش قرار می‌گیرد. ویژگی شماره ۱ میانگین خوشه‌ها و ویژگی شماره ۲ آنتروپی خوشه‌ها می‌باشد.

در ویژگی‌های شماره ۲۰ و ۲۱ بر اساس معیار شباهت کسینوسی (رابطه (۳-۴))، بین نوبت‌ها خوشه-بندی انجام می‌گیرد. شباهت کسینوسی نوبت جدید (نوبتی که می‌خواهد خوشه‌بندی شود) با تمام نوبت‌های قبل از خودش محاسبه می‌شود. برای محاسبه شباهت کسینوسی دو نوبت، باید از بردارهای یکسانی برای نمایش هر نوبت استفاده کرد. کلمه‌های نوبت جدید در ابتدا و بعد از آن کلمه‌های نوبت دوم می‌آید. این بردار برای هر دو نوبت بر اساس میزان تکرار کلمه‌ها در هر نوبت پر و سپس بین دو بردار شباهت کسینوسی محاسبه می‌شود. در این معیار می‌خواهیم خوشه‌بندی به‌گونه‌ای باشد که طول دو نوبت در خوشه‌بندی تأثیر داشته باشد. در واقع به‌گونه‌ای نباشد که دو نوبت با تعداد کلمه

زیاد به خاطر یکسان بودن فقط یک کلمه، در یک خوشه قرار گیرند. به همین دلیل از یک مقدار حد آستانه برای نزدیک بودن دو نوبت به هم استفاده شده است. نوبت جدید به هر نوبت قبلی که نزدیک‌تر و شباهت کسینوسی آن‌ها بیشتر از θ باشد، به خوشه‌ی همان نوبت تعلق می‌گیرد. اگر یک نوبت به هیچ خوشه‌ای نزدیک نبود دو حالت به وجود می‌آید: ۱-خوشه‌ای که حداقل ٪۵۰ نوبت‌هایش شباهت کسینوسی کمتر از θ دارد به آن خوشه تعلق می‌گیرد. ۲-اگر به هیچ خوشه‌ای با این ویژگی‌ها نزدیک نبود خودش جداگانه خوشه جدید تشکیل می‌دهد.

ویژگی شماره ۲۰ میانگین خوشه‌ها و ویژگی شماره ۲۱ آنتروپی خوشه‌ها می‌باشند.

$$cosine(s_i . s_j) = \frac{\overrightarrow{s_i} \cdot \overrightarrow{s_i}}{\|\overrightarrow{s_i}\| \cdot \|\overrightarrow{s_j}\|} \qquad (۴-۳)$$

ویژگی‌های شماره ۲۲ و ۲۳ میانگین و آنتروپی خوشه‌هایی هستند که مکالمه بر اساس معیار شباهت رابطه (۳-۵) خوشه‌بندی شده باشد. مقاله‌ی [۴۰] از رابطه (۳-۵) برای اندازه‌گیری انسجام و شباهت بین جمله‌های متن استفاده کرده است. در اینجا برای اندازه‌گیری انسجام در مکالمه چون جمله‌های مکالمه کوتاه است، شباهت را بین یک نوبت و کیسه‌ای از کلمه‌های هر خوشه محاسبه می‌کنیم. به هر خوشه‌ای شباهت بیشتری داشت به همان خوشه تعلق دارد. تفاوت این ویژگی با ویژگی‌های شماره ۱و ۲ که بر اساس تکرار کلمه‌ها خوشه‌بندی انجام می‌شود در این است که با تقسیم بر جمع طول جمله و طول کیسه کلمه‌ها، میزان نزدیکی به خوشه‌ها نرمال می‌شود.

$$\mathrm{Sim}\left(\mathrm{BOW_j} . \mathrm{S_i}\right) = \frac{2|\mathrm{words(S_i)} \cap \mathrm{words(BOW_j)}|}{|\mathrm{words(S_i)}| + |\mathrm{words(BOW_j)}|} \qquad (۵-۳)$$

در رابطه (۳-۵) $\mathrm{BOW_j}$ کیسه کلمه‌های خوشه‌ی j ام و $\mathrm{S_i}$ جمله i ام از مکالمه است.

۳-۳-۳- میزان تکرار کلمه‌ها در مکالمه

این ویژگی‌ها تعداد کلمه‌های تکرار شده در مکالمه را به صورت مختلف محاسبه می‌کنند. یکی از ویژگی‌ها، تعداد نوبت‌هایی است که بیش از مقدار مشخصی کلمه تکرار شده با تاریخچه مکالمه دارند.

تعداد نوبت‌ها شمارش و با تقسیم بر تعداد کل نوبت‌های مکالمه نرمال می‌شوند. حداقل تعداد تکرار به روش آزمون و خطا و یا با روش‌های امتیازدهی به ویژگی‌ها محاسبه می‌شود. ۳ عدد بهتر به عنوان ویژگی‌های ۳، ۴ و ۵ در نظر گرفته می‌شوند.

ویژگی‌های دیگری که می‌تواند در تعیین میزان انسجام مؤثر باشد، میانگین و واریانس تعداد کلمه‌های تکرار شده می‌باشد. در هر نوبت تعداد کلمه‌های مشابه با تاریخچه گفتگو محاسبه و تقسیم بر تعداد کل کلمه‌های آن نوبت نرمال می‌شود. میانگین و واریانس این اعداد در هر مکالمه می‌تواند معیاری برای ارزیابی انسجام باشد. ویژگی ۶ و ۷ به ترتیب مربوط به میانگین و واریانس این اعداد می‌باشند.

۳-۳-۴- بیشترین تعداد تکرار کلمه‌ها

ویژگی‌های ۸ تا ۱۵ مربوط به بیشترین تعداد کلمه‌های تکراری در کل مکالمه می‌باشند. در دو حالت، تعداد کلمه‌های تکراری نرمال می‌شوند، یکی با تقسیم بر تعداد کل کلمه‌ها و دیگری تقسیم بر تعداد کلمه‌های منحصر به فرد مکالمه. ویژگی‌های شماره ۸ تا ۱۱ و ۱۲ تا ۱۵ تعداد تکرار ۴ کلمه پرتکرار در مکالمه می‌باشند که به ترتیب با تقسیم بر تعداد کل کلمه‌ها و تعداد کلمه‌های منحصر به فرد نرمال می‌شوند.

۳-۳-۵ تکرار کلمه با توجه به طول مکالمه

تعداد تکرار کلمه‌ها در هر نوبت مکالمه با توجه به فاصله تکرار یا فاصله دو نوبت که کلمه در آن‌ها تکرار شده‌اند، ارزش متفاوتی دارند. هرچه فاصله دو نوبت مکالمه بیشتر، ارزش تکرار آن کلمه کمتر است. همچنین هرچه کلمه‌ی تکرار شده در دو نوبت به هم نزدیک‌تر یعنی هر چه فاصله دو نوبت که کلمه در آن‌ها تکرار شده کمتر باشد، ارزش بیشتری دارد. در ویژگی شماره ۱۶ بر اساس همین تعریف میانگین تعداد تکرار هر نوبت با نوبت‌های قبلی با میزان ضریب β محاسبه شده است. در محاسبه این ویژگی ارزش تعداد تکرار هر نوبت با نوبت قبلی‌اش β و با کل نوبت‌های گذشته به غیر از نوبت قبلی $\beta - ۱$ می‌باشد. این ویژگی به صورت بازگشتی برای هر نوبت محاسبه می‌شود. سپس بین تمام اعداد

به دست آمده برای تمام نوبت‌ها میانگین می‌گیریم. برای تأثیر طول نوبت‌های مکالمه تعداد تکرار هر دو نوبت با تقسیم بر اندازه‌ی کوچک‌ترین نوبت نرمال می‌شوند. در رابطه‌های (۳-۶)، (۳-۷)، (۳-۸) و (۳-۹) نحوه‌ی محاسبه این ویژگی آمده است.

$$x(i.j) = \frac{|\,words(s_i) \cap words(s_j)|}{min\{|\,words\,(s_i)\,|\,.\,|words(s_j)|\}} \qquad (۳-۶)$$

$$f_1(i) = \beta x(i.i-1) + (1-\beta) f_2(i.i-2) \qquad (۳-۷)$$

$$f_2(i.j) = \begin{cases} \beta x(i.j) + (1-\beta) f_2(i.j-1) & j \geq 0 \\ 0 & j < 0 \end{cases} \qquad (۳-۸)$$

$$m_{16} = \frac{\sum_{i=0}^{n} f_1(i)}{n} \qquad (۳-۹)$$

در رابطه (۳-۶) تعداد کلمه‌های تکراری دو نوبت از مکالمه (S_i و S_j) محاسبه و با تقسیم بر طول کمترین نوبت، تعداد تکرارها نرمال می‌شود. در رابطه‌های (۳-۷) و (۳-۸) تعداد تکرار کلمه‌های نوبت iام با توجه به عمق نوبتی که کلمه‌ها در آن تکرار شده‌اند محاسبه می‌شود. در انتها بین تمام اعداد به دست آمده برای تمام نوبت‌ها میانگین گرفته می‌شود(رابطه (۳-۹)). n تعداد نوبت‌های مکالمه می‌باشد.

ویژگی شماره ۱۷ مانند ویژگی شماره ۱۶ محاسبه می‌شود اما مقدار β و $1-\beta$، ۱ در نظر گرفته شده است. یعنی عمق نوبت‌ها در محاسبه تأثیر داده نشده است و. در هر نوبت بین تعداد تکرارها با نوبت‌های قبلی و در آخر بین نتایج نوبت‌ها میانگین گرفته می‌شود. تفاوت این ویژگی با ویژگی شماره ۶ در تأثیر طول نوبت‌های تاریخچه گفتگو می‌باشد.

۳-۳-۶- تکرار کلمه در دو نوبت متوالی

تاکنون ویژگی‌های استخراج شده بر اساس تکرار کلمه‌ها در کل مکالمه و یا تمام نوبت‌های گذشته است. در ویژگی‌های ۱۸ و ۱۹ تعداد تکرارها فقط در دو نوبت متوالی محاسبه می‌شود.

ویژگی شماره ۱۸، یکی از ویژگی‌هایی می‌باشد که مقاله‌ی [۴۰] برای محاسبه انسجام متن از آن استفاده کرده است. در این ویژگی میانگین تعداد کلمه‌های تکراری دو نوبت متوالی محاسبه می‌شود. تعداد کلمه‌های تکراری هر دو نوبت متوالی شمارش و با تقسیم بر جمع طول هر دو نوبت نرمال می-شود.

$$sim(s_i.s_{i+1}) = \frac{2|words(s_i) \cap words(s_{i+1})|}{|words(s_i)|+|words(s_{i+1})|} \qquad (۱۰-۳)$$

$$M_{18} = \frac{\sum_{i=1}^{n-1} sim(s_i.s_{i+1})}{n-1} \qquad (۱۱-۳)$$

همان‌طور که در رابطه (۳-۱۰) نشان داده شده است، شباهت هر دو نوبت متوالی با استفاده از تعداد کلمه‌های تکراری آن‌ها محاسبه و با تقسیم بر جمع طول دو نوبت نرمال می‌شود. در رابطه (۳-۱۱) بین شباهت‌های هر دو نوبت در کل مکالمه، میانگین گرفته می‌شود. n تعداد کل نوبت‌های مکالمه می‌باشد.

در ویژگی ۱۹ شباهت دیگری برای میزان شباهت دو نوبت متوالی تکرارها استفاده شده است. مقاله‌ی [۹] برای تولید خلاصه یک متن، متن را یک گراف نمایش داده است. جمله‌ها نودهای گراف و برای یال‌ها اعدادی بر اساس انسجام بین جمله‌ها نسبت داده است. یال‌های گراف را با سه روش مقداردهی کرده است. یکی از روش‌ها شباهت کسینوسی بین هر دو جمله بوده است. مشابه این روش در ویژگی شماره ۱۹ هر نوبت را با برداری از کلمه‌ها و تعداد تکرار آن‌ها نشان داده شده است. سپس بین هر دو نوبت شباهت کسینوسی (رابطه (۳-۴)) را محاسبه و بین نتایج به دست آمده مانند رابطه (۳-۱۱) میانگین گرفته شده است.

۳-۴- استفاده از همنشینی کلمه‌ها

تاکنون ویژگی‌هایی بر پایه تکرار کلمه‌های کلیدی معرفی شده است. فقط با استفاده از نحوه‌ی تکرار کلمه‌های کلیدی در یک مکالمه نمی‌توان به میزان انسجام آن‌ها پی برد. ممکن است یک مکالمه‌ی

منسجم هیچ تکرار کلمه‌ای در آن رخ ندهد و کلمه‌ها از نظر معنا به هم مرتبط باشند. با استفاده از یک پایگاه داده‌ی همنشینی بین کلمه‌ها، می‌توان تشخیص صحیح میزان انسجام مکالمه را افزایش داد. پایگاه داده‌ی همنیشینی شامل مجموعه‌ای از کلمه‌ها و میزان همنشینی آن‌ها می‌باشد. اندازه‌ی همنشینی به معنای میزان عددی است که دو کلمه می‌توانند باهم در یک جمله، پاراگراف و یا کل متن حضور داشته باشند. با استفاده از چنین پایگاه داده‌ای می‌توان دقیق‌تر یک متن را مورد بررسی قرار داد.

ویژگی‌های معرفی شده برای اندازه‌گیری انسجام در این بخش، تعدادی از ویژگی‌های گذشته است که باهمنشینی ترکیب شده‌اند و همچنین دارای چند ویژگی جدید می‌باشند.

۳-۴-۱- خوشه‌بندی

خوشه‌بندی مکالمه در بخش ۳-۳-۱ توضیح داده شده است. معیار شباهت دیگری که برای خوشه-بندی استفاده شده، بر اساس میزان همنشینی کلمه‌ها می‌باشد. میزان همنشینی کلمه‌های موجود در نوبت‌ها در خوشه‌بندی آن‌ها مؤثر می‌باشد. با استفاده از پایگاه داده، میزان همنشینی کلمه‌ها را محاسبه می‌کنیم. نوبت‌ها به خوشه‌هایی تعلق می‌گیرند که بیشترین میانگین همنشینی بین کلمه-های موجود در آن‌ها و خوشه‌ها وجود داشته باشد. می‌توان یک حد آستانه برای میانگین همنشینی نوبت‌ها در خوشه‌بندی تعیین کرد، زیرا تمام کلمه‌های موجود حداقل مقدار کمی همنشین می‌باشند. با تعیین حد آستانه برای ترکیب، نوبت‌های دارای میانگین همنشینی کم باهم ترکیب نمی‌شوند. ویژگی‌های شماره ۲۶ و ۲۷ به ترتیب میانگین و آنتروپی خوشه‌ها در این معیار خوشه‌بندی می‌باشند.

۳-۴-۲- تعداد کلمه‌های همنشین با تاریخچه مکالمه

همان‌طور که گفته شد علاوه بر تکرار، همنشینی کلمه‌ها با تاریخچه گفتگو نیز در میزان انسجام مکالمه مؤثر می‌باشد. ویژگی‌های ۳۱ تا ۳۳ مشابه ویژگی‌های شماره ۳ تا ۵ می‌باشند. این ویژگی‌ها میانگین تعداد نوبت‌هایی می‌باشند که بیشتر از یک مقدار مشخص، کلمه‌های همنشین با تاریخچه

گفت گو داشته باشند. اگر کمترین اندازه همنشینی صفر و بیشترین ۱ باشد، اندازه همنشینی دو کلمه هر عددی بین ۰ تا ۱ می‌تواند باشد. با تعیین حداقل اندازه برای همنشینی دو کلمه می‌توان دقت اندازه‌گیری انسجام را افزایش داد.

ویژگی‌های ۳۱، ۳۲ و ۳۳ در حالت همنشینی به ترتیب مشابه ویژگی‌های ۳ تا ۵ در بخش تکرار کلمه‌های کلیدی می‌باشند. ویژگی‌های ۲۹ و ۳۰ به ترتیب میانگین و واریانس کلمه‌های همنشین با تاریخچه گفتگو مانند ویژگی‌های شماره ۶ و ۷ می‌باشند.

۳-۴-۳- میانگین همنشینی

اندازه‌ی میانگین شباهت هر نوبت با نوبت‌های گذشته می‌تواند یک معیار برای تشخیص انسجام باشد. در ویژگی شماره ۲۴ و ۲۸ بین هر دو خط متوالی میانگین همنشینی کلمه‌ها محاسبه می‌شود. محاسبه میانگین همنشینی بین دو خط متوالی در رابطه (۳-۱۲) نشان داده شده است. در این رابطه تابع co میزان همنشینی دو کلمه را محاسبه می‌کند. $W_{s_{n.i}}$ مربوط به‌کلمه iام از جمله‌ی نوبت sام می‌باشد. ویژگی شماره ۲۴ به‌طور میانگین در چند نوبت این میزان همنشینی (M_n) بیشتر از θ و در شماره ۲۸ میانگین همنشینی‌های دو خط متوالی در کل مکالمه محاسبه شده است.

$$M_n = \frac{\sum_i \sum_j co(W_{s_n.i}.W_{s_{n+1}.j})}{|s_n|+|s_{n+1}|} \qquad (۱۲-۳)$$

ویژگی شماره ۲۵ به‌طور میانگین، میانگین همنشینی هر نوبت با تمام نوبت‌های گذشته‌اش چه اندازه می‌باشد(رابطه (۳-۱۳)). در رابطه (۳-۱۳) میانگین شباهت هر نوبت با نوبت‌های گذشته‌اش محاسبه می‌شود. ویژگی شماره ۲۵ میانگین بین مقدار تمام نوبت‌ها می‌باشد.

$$Ml_n = \frac{\sum_{y=2}^{n} \sum_{x=y-1}^{1} \frac{\sum_i \sum_j co(W_{s_y.i}.W_{s_x.j})}{|s_y|+|s_x|}}{n-1} \qquad (۱۳-۳)$$

۳-۴-۴- شناسایی شکست‌های مکالمه

همان‌طور که گفته شد برای ساخت یک مکالمه با انسجام کمتر، تعدادی از نوبت‌ها تغییر می‌کنند. در واقع در مکالمه شکست ایجاد شده است. با شناسایی نوبت‌هایی که باعث شکست در مکالمه شده‌اند، می‌توان به میزان انسجام پی برد. از نوبت‌هایی که در آن‌ها شکست رخ داده با عنوان شکست، ویژگی-هایی استخراج می‌شود. ویژگی‌ها از همان نوبت در مکالمه اصلی (مکالمه با انسجام عالی) با عنوان اتصال استخراج می‌شود. به این صورت مجموعه‌ای از ویژگی‌ها داریم که به دو دسته اتصال و شکست تقسیم شده‌اند. یکی از الگوریتم‌های دسته‌بندی با این داده‌ها آموزش می‌بینند. از نتایج دسته‌بندی می‌توان در شناسایی نوبت‌های شکسته شده استفاده کرد. در ادامه ویژگی‌های معرفی شده در شناسایی شکست و اتصال‌ها و نحوه استفاده از آن‌ها به عنوان ویژگی در تعیین اندازه انسجام آمده است.

ویژگی‌های استفاده شده برای شناسایی شکست و اتصال‌ها بر اساس میزان همنشینی کلمه‌ها می‌باشد. از ویژگی‌های شماره ۱۶، ۱۷و ۲۵ و سه ویژگی جدید استفاده شده است. در استفاده از ویژگی‌های ۱۶ و ۱۷ علاوه بر تکرار، تعداد کلمه‌هایی که همنشینی آن‌ها نیز از یک حد مشخصی بیشتر باشند، شمارش می‌شوند.

اولین ویژگی از سه ویژگی جدید استفاده شده، میانگین تعداد کلمه‌هایی است که همنشینی آن‌ها با تاریخچه از یک حد مشخص بیشتر می‌باشد. ویژگی دوم، میانگین اندازه بیشترین همنشینی کلمه‌های نوبتی که می‌خواهد دسته‌بندی شود با نوبت‌های گذشته می‌باشد. ویژگی آخر میانگین تعداد نوبت‌هایی است که حداقل یک کلمه همنشنین با نوبتی که می‌خواهد دسته‌بندی شوند داشته باشند. در تمام ویژگی‌ها اگر میزان همنشینی از یک مقدار مشخصی بیشتر باشد آن کلمه‌ها شمارش می-شوند. این مقدار مشخص با سعی و خطا به دست می‌آید.

ویژگی‌های توضیح داده شده برای هر نوبت شکست و اتصال همان نوبت، محاسبه می‌شود. یکی از الگوریتم‌های دسته‌بندی با داده‌های به دست آمده، برای دسته‌بندی به دو دسته شکست و اتصال آموزش داده می‌شود.

برای شناسایی شکست‌ها در یک مکالمه، مکالمه بر اساس مکان‌های شکست دسته‌بندی می‌شود. نوبت اول مکالمه دارای شکست نمی‌باشد و دسته‌ی اصلی را تشکیل می‌دهد. از نوبت دوم تا انتهای مکالمه هر نوبت به ترتیب با استفاده از الگوریتم دسته‌بندی به هر دسته‌ای که تعلق داشت داخل مجموعه‌ی آن دسته می‌رود. تعلق داشتن به آن دسته به معنای پیش‌بینی اتصال با آن دسته توسط الگوریتم دسته‌بندی می‌باشد. اگر به چند دسته تعلق داشت، دو دسته باهم ترکیب می‌شوند و اگر به هیچ دسته‌ای تعلق نداشت (پیش‌بینی شکست)، دسته‌ای جدید تشکیل می‌شود. بعد از دسته‌بندی یک مکالمه از دسته‌های ایجاد شده، ۳ ویژگی برای تعیین سطح انسجام محاسبه می‌شود. ویژگی شماره ۳۴ تعداد نوبت‌هایی است که در دسته‌ها به تنهایی قرار گرفته‌اند. در واقع تعداد نوبت‌هایی که به عنوان شکست پیش‌بینی شده‌اند، با تقسیم بر تعداد طول مکالمه نرمال می‌شوند. با این ویژگی درواقع میزان انسجام مکالمه پیش‌بینی می‌شود. ویژگی‌های ۳۵ و ۳۶ به ترتیب میانگین و آنتروپی دسته‌ها می‌باشند. همان‌طور که در خوشه‌بندی گفته شد میانگین و آنتروپی خوشه‌ها در میزان انسجام مؤثر هستند.

۳-۴- دسته‌بندی مکالمه‌ها

ویژگی‌های معرفی شده در دسته‌بندی مکالمه‌ها برحسب انسجام، مؤثر می‌باشند. برای اثبات و یا رد این موضوع لازم است با الگوریتم‌های دسته‌بندی، مکالمه‌ها دسته‌بندی شوند. سپس از روی نتایج دسته‌بندی، میزان کارایی ویژگی‌ها مورد ارزیابی قرار می‌گیرند. به دلیل آنکه الگوریتم‌های مختلف دسته‌بندی نتایج متفاوتی از دسته‌بندی مکالمه دارند، بهتر است با چند الگوریتم دسته‌بندی شوند و بهترین آن‌ها به عنوان دسته‌بند در این نوع مکالمه‌ها انتخاب شود. روش‌های دسته‌بندی مورد استفاده

در این پژوهش بیزین ساده[1]، ماشین بردار پشتیبان[2] و درخت تصمیم[3] می‌باشند. هر کدام از روش‌های دسته‌بندی با استفاده از نرم‌افزار وکا پیاده‌سازی شده و تنظیمات پیش‌فرض وکا می‌باشد.

۳-۴-۱- بیزین ساده

کلاس‌بندی بیزین ساده به دلیل اینکه از نظر هزینه محاسباتی به پردازنده و حافظه کمی نیاز دارد، بسیار کارآمد است. همچنین این روش نیاز به مقدار کمی از داده‌ها برای آموزش دارد و زمان آموزش نسبت به سایر روش‌ها کمتر است.

ایده اصلی بیزین ساده، پیدا کردن احتمال دسته‌های متنی با استفاده از احتمال‌های اتصال ویژگی و دسته‌ها است. احتمال ویژگی X و دسته C برابر است با:

$$P(C|X) = \frac{P(X|C)\times P(C)}{P(X)} \qquad (۱۴-۳)$$

به بیانی دیگر بیزین ساده یک طبقه‌بندی تولیدی[4] است که احتمال‌های پیشین P(X | C) و P(C) را از داده‌های آموزشی برآورد می‌کند و احتمال پسین P(C | X) را بر اساس احتمالات پیشین تولید می‌کند.

۳-۴-۲-ماشین بردار پشتیبان

ماشین بردار پشتیبان، یک طبقه‌بندی جدا کننده[5] است که هیچ پیش‌فرضی را بر اساس داده‌های آموزشی ایجاد نمی‌کند و به‌طور مستقیم P(C | X) را تخمین می زند.

در این پایان‌نامه از تابع هسته چندجمله‌ای نرمال شده برای آزمایش ماشین بردار پشتیبان استفاده شده است.

[1] Naive Bayes
[2] Support Vector Machin(SVM)
[3] Decision Tree(DT)
[4] Generative Classifier
[5] Discriminative Classifier

۳-۴-۳- درخت تصمیم

درخت تصمیم روشی برای نمایش یک مجموعه از قوانین است که منتهی به یک دسته یا مقدار می-شوند. درخت تصمیم از طریق جداسازی متوالی داده‌ها به گروه‌های مجزا ساخته می‌شود و هدف در این فرایند، افزایش فاصله بین گروه‌ها در هر جداسازی است. این روش ساختار ساده و قابل فهمی دارد و با ویژگی‌های پیش‌بینی کننده به‌خوبی کار می‌کند. اگر به درخت اجازه داده شود که بدون محدویت رشد کند، علاوه بر افزایش زمان ساخت، ممکن است داده‌ها بیش‌برازش[1] شوند. برای جلوگیری از بیش برازش دو روش وجود دارد. در روش اول باید قبل از رسیدن به مرحله‌ای که داده-های آموزشی به‌طور کامل دسته‌بندی شوند، رشد درخت را متوقف نمود. در روش دوم به درخت اجازه داده می‌شود که کامل رشد کند و سپس شاخه‌هایی که مفید نیستند هرس می‌شوند. در این پژوهش برای جلوگیری از بیش برازش از C4.5 استفاده شده است که توسط معیار اطمینان[2] ۰/۲۵ شاخه‌ها را هرس می‌کند و حداقل تعداد هر نمونه در هر برگ برابر ۲ است.

۳-۵- جمع‌بندی

در این فصل ابتدا تولید پایگاه داده‌ای از مکالمه‌ها با انسجام‌های مختلف توضیح داده شده است. برای مکالمه‌ها از مکالمه‌های آموزش زبان انگلیسی استفاده شده است. این مکالمه‌ها انسجام عالی دارند و با ترکیب و جانشین کردن نوبت‌ها، مکالمه با انسجام کمتر ساخته می‌شود. پایگاه داده‌ی تولید شده شامل ۳ دسته مکالمه با انسجام عالی، متوسط و ضعیف می‌باشد. هر دسته شامل ۷۰۰ مکالمه برای آموزش و ۱۰۰ مکالمه برای بخش آزمون می‌باشد.

میزان سطح انسجام مکالمه با استفاده از ویژگی‌هایی که از مکالمه استخراج می‌شود، قابل پیش‌بینی می‌باشد و ۳۶ ویژگی که در میزان انسجام مکالمه مؤثر می‌باشند، پیشنهاد داده شده است.

[1] Overfitting
[2] Confidence Factor

فصل چهارم:

نتایج و آزمایش‌ها

۴-۱- مقدمه

در این فصل ویژگی‌های معرفی شده مورد ارزیابی قرار می‌گیرند. ابتدا برای ویژگی‌هایی که مقادیر ثابت دارند، با استفاده از مجموعه آموزش و الگوریتم‌های انتخاب ویژگی بهترین مقدار آن‌ها به دست می‌آیند. در گام بعدی از بین تمام ویژگی‌های معرفی شده، زیر مجموعه‌ای از بهترین ویژگی‌ها برای ارزیابی انتخاب می‌شوند. ویژگی‌های استخراج شده از مکالمه‌ها با انواع الگوریتم‌های دسته‌بندی، به سه دسته عالی، متوسط و ضعیف از نظر انسجام تقسیم می‌شوند. میزان دقت این الگوریتم‌ها نشان از کارایی ویژگی‌ها در تعیین انسجام مکالمه می‌باشند. در انتها به آنالیز نتایج الگوریتم‌ها و علت جواب‌ها پرداخته می‌شود.

۴-۲- انتخاب ویژگی

فصل گذشته شامل تولید پایگاه داده‌ای از مکالمه‌ها و معرفی ویژگی‌های پیشنهادی برای ارزیابی انسجام مکالمه بوده است. ویژگی‌ها به دو بخش تکرار کلمه‌های کلیدی و همنشینی آن‌ها تقسیم شده‌اند. برای به دست آوردن مقدار تعدادی از ویژگی‌های معرفی شده نیاز به مقدارهای ثابتی می-باشد. این مقدارهای ثابت باید به‌گونه‌ای انتخاب شوند که بیشترین تأثیر را در تعیین سطح انسجام داشته باشند. پس نیاز است به روش آزمون و خطا و یا با الگوریتم‌های انتخاب ویژگی بهترین مقادیر انتخاب شوند. روش آزمون و خطا یعنی با مقادیر مختلف ویژگی‌ها استخراج و سپس ویژگی که کمترین خطا را در دسته‌بندی دارد، به عنوان ویژگی مناسب در نظر گرفته شود. بعد از مشخص شدن تمام ویژگی‌ها باید از بین آن‌ها بهترین ویژگی‌ها انتخاب شوند. در مجموع ۲۴ ویژگی مربوط به تکرار کلمه‌ها برای تعیین سطح انسجام مکالمه پیشنهاد شده است. ویژگی‌ها برحسب تعداد تکرارهای کلمه در حالت‌های مختلف و خوشه‌بندی با معیارهای متفاوت می‌باشند. ممکن است تعدادی از ویژگی‌های پیشنهاد شده به هم وابسته باشند. یعنی با افزایش یکی از آن‌ها ویژگی دیگر افزایش یا کاهش یابد. ویژگی‌هایی که به هم وابسته هستند ویژگی‌های اضافی می‌باشند. همچنین تعدادی از ویژگی‌ها به

عنوان ویژگی‌های اضافی در تعیین سطح انسجام تأثیر زیادی ندارند. انتخاب زیر مجموعه‌ای از بهترین ویژگی‌ها در بالا بردن تشخیص صحیح الگوریتم‌های دسته‌بندی بسیار مؤثر می‌باشد.

۴-۲-۱ مقداردهی مقادیر ثابت

همان‌طور که گفته شد بعضی از ویژگی‌های معرفی شده دارای مقادیر ثابتی هستند. این مقادیر با توجه به پایگاه داده، بهترین مقداری که می‌تواند در دسته‌بندی مکالمه مفید باشند هستند. برای به دست آوردن بهترین مقدار ثابت از آزمون آماری انتخاب ویژگی به نام بهره اطلاعات[1] استفاده می‌شود (رابطه (۴-۱)).

$$\text{InfoGain}(\text{Class}.\text{Attribute}) = \text{H}(\text{Class}) - \text{H}(\text{Class} \mid \text{Attribute}) \quad (۴-۱)$$

ویژگی‌ها با مقادیر ثابت مختلف از نمونه مکالمه‌ها (مجموعه آموزش) استخراج می‌شوند. بهره اطلاعات مشخص می‌کند کدام ویژگی‌ها به تنهایی بهتر قادر خواهند بود نمونه مکالمه‌ها را دسته‌بندی کنند. در هر ویژگی مقدار ثابتی که بهره اطلاعاتش بیشتر باشد، به عنوان ثابت انتخاب می‌شود. بهره اطلاعات با استفاده از نرم افزار وکا اجرا شده است. مقادیر ویژگی‌ها به شرح زیر می‌باشند.

ویژگی‌های شماره ۳، ۴ و ۵: در بخش ویژگی بر اساس میزان تکرار کلمه‌ها گفته شد، تعداد نوبت‌هایی که کلمه‌های تکراری با سابقه‌ی گفتگوی آن‌ها از مقدار مشخصی (β) بیشتر باشند شمارش می‌شوند. این ویژگی با مقدار حداقل‌های مختلف به دست آورده شده است. بهترین مقادیر از نظر بهره اطلاعات به این گونه می‌باشند.

ویژگی شماره ۳، میانگین تعداد نوبت‌هایی که حداقل ٪۱۰ و ویژگی شماره ۴، میانگین تعداد نوبت‌هایی که حداکثر ٪۱۰ و حداقل ٪۵ از کلمه‌های تاریخچه مکالمه را داشته می‌باشند (جدول (۴-۱)).

[1] Information Gain

جدول (۴-۱): بهره اطلاعات ویژگی‌های شماره ۳ و ۴ با حداقل β

β	۵	۱۰	۲۰	۲۵	۳۰
بهره اطلاعات	۰/۲۳	۰/۲۶	۰/۲۰	۰/۲۰	۰/۱۹

همان‌طور که در جدول (۱-۴) نشان داده شده است با توجه به کم بودن تفاوت بهره اطلاعات بین مقدارهای ثابت ۵ و ۱۰ درصد، ویژگی ۴ام بین این دو مقدار در نظر گرفته شده است.

ویژگی ۵ام، تعداد نوبت‌هایی را که بیشتر از ٪۴۰ کلمه‌های نوبت با تاریخچه‌ی گفتگو تکراری باشد بیان می‌کند (جدول (۴-۲)).

جدول (۴-۲): بهره اطلاعات ویژگی شماره ۵ با حداقل β

β	۱۰	۲۰	۲۵	۳۰	۳۵	۴۰	۴۵	۵۰	۵۵
بهره اطلاعات	۰/۲۴	۰/۲۴	۰/۲۴	۰/۲۵	۰/۲۵	۰/۲۷	۰/۲۶	۰/۲۶	۰/۲۵

ویژگی شماره ۱۶: در این ویژگی ارزش تکرار کلمه با توجه به عمق مکالمه کاهش می‌یابد. مقدار این ویژگی از مقدار ۰ تا ۰/۹ β با گام ۰/۱ام استخراج شده است(جدول(۴-۳)). بهترین مقدار β که ارزش اطلاعات آن از همه بیشتر است ۰/۵ می‌باشد.

جدول (۴-۳): بهره اطلاعات میانگین تکرار کلمه‌ها باارزش β نسبت به عمق تکرار

β	۰/۱	۰/۲	۰/۳	۰/۴	۰/۵	۰/۶	۰/۷	۰/۸	۰/۹
بهره اطلاعات	۰/۱۳	۰/۱۹	۰/۲۰	۰/۲۰	۰/۲۰	۰/۲۰	۰/۱۹	۰/۱۹	۰/۱۹

ویژگی شماره ۲۰ و ۲۱: این دو ویژگی به ترتیب میانگین و آنتروپی خوشه‌ها در خوشه‌بندی با معیار کسینوسی هستند. همان‌طور که گفته شد تعیین خوشه به‌گونه‌ای باید باشد که طول دو نوبت در میزان تکرار کلمه‌ها مؤثر باشد. یعنی اگر طول دو نوبت زیاد باشد، تعداد بیشتری کلمه‌ی مشابه باید باهم داشته باشند تا در یک خوشه قرار گیرند. به همین دلیل یک حد آستانه (θ) برای نزدیک دو نوبت بودن تعریف شده است. میانگین و آنتروپی خوشه‌ها با θ متفاوت از مجموعه آموزش استخراج شده است(جدول (۴-۴) و (۴-۵)). θ از ۰ تا ۰/۹ با گام ۰/۱ تغییر کرده است. بهترین بهره اطلاعات مربوط به حد آستانه ۰/۳ می‌باشد.

جدول (۴-۴): بهره اطلاعات آنتروپی خوشه‌ها در خوشه‌بندی با معیار کسینوسی

θ	۰/۱	۰/۲	۰/۳	۰/۴	۰/۵	۰/۶	۰/۷	۰/۸	۰/۹
بهره اطلاعات	۰/۲۲	۰/۲۵	۰/۲۶	۰/۲۱	۰/۱۶	۰/۲۲	۰/۲۲	۰/۲۲	۰/۲۲

جدول (۴-۵): بهره اطلاعات میانگین خوشه‌ها در خوشه‌بندی با معیار کسینوسی

θ	۰/۱	۰/۲	۰/۳	۰/۴	۰/۵	۰/۶	۰/۷	۰/۸	۰/۹
بهره اطلاعات	۰/۲۱	۰/۲۳	۰/۲۵	۰/۲۴	۰/۱۹	۰/۲۱	۰/۲۱	۰/۲۱	۰/۲۱

۴-۲-۲- انتخاب بهترین ویژگی‌ها

در دقت دسته‌بندی، انتخاب بهترین زیر مجموعه از ویژگی‌ها بسیار مهم است. در این بخش با استفاده از الگوریتم‌های انتخاب ویژگی، زیر مجموعه‌ای از بهترین ویژگی‌ها انتخاب می‌شوند. طبق مقاله‌ی [۴۱] ویژگی‌هایی مناسب هستند که همبستگی[1] زیادی با کلاس و همبستگی کمی با دیگر ویژگی‌ها

[1] Correlation

داشته باشند. با استفاده از نرم افزار وکا از الگوریتم جستجوی اولین بهترین انتخاب[1] استفاده شده تا فضای ویژگی‌ها را جستجو و بهترین زیرمجوعه از ویژگی‌ها را انتخاب کند. انتخاب بهترین زیرمجموعه از ویژگی‌ها با استفاده از ارزیابی همبستگی بین ویژگی‌ها (CFS)[2] که کمترین همبستگی را باهم و بیشترین همبستگی را با کلاس داشته باشند(رابطه (۴-۱))، انتخاب می‌شوند. در رابطه (۴-۱) r_{cf} ضریب همبستگی پیرسون[3] بین کلاس و ویژگی‌های زیر مجموعه و $r_{f_1f_2}$ ضریب همبستگی پیرسون بین دو ویژگی زیرمجموعه می‌باشد. ضریب همبستگی پیرسون حاصل تقسیم کواریانس[4] و ضرب انحراف معیار[5] دو ویژگی می‌باشد.

$$CFS = max_{s_k}\left[\frac{r_{cf_1}+r_{cf_2}\dots+r_{cf_k}}{\sqrt{k+2(r_{f_1f_2}+\dots+r_{f_1f_k}\dots+r_{f_kf_{k-1}})}}\right] \qquad (۴-۱)$$

ویژگی‌های انتخاب شده شماره‌های ۱، ۲، ۴، ۶، ۷، ۹، ۱۱، ۱۲، ۱۳، ۱۴، ۱۸، ۱۹، ۲۱ و ۲۳ می-باشند.

ویژگی ۱: میانگین تعداد خوشه‌ها در خوشه‌بندی بر اساس بیشترین تعداد تکرار کلمه‌ها.

ویژگی ۲: آنتروپی خوشه‌ها در خوشه‌بندی بر اساس بیشترین تعداد تکرار کلمه‌ها.

ویژگی ۴: میانگین تعداد نوبت‌هایی که با حداقل ۵ و حداکثر ۱۰ درصد از کلمه‌های موجود در تاریخچه گفتگو تکراری باشند.

ویژگی ۶: میانگین تعداد نوبت‌هایی که حداقل یک کلمه تکراری با تاریخچه گفت گو دارند.

ویژگی ۷: واریانس تعداد نوبت‌هایی که حداقل یک کلمه تکراری با تاریخچه گفت گو دارند.

[1] Best-First Search
[2] Correlation Feature Selection
[3] Pearson's Correlation Coefficient
[4] Covariance
[5] Standard deviation

ویژگی‌های ۹ و ۱۱: به ترتیب دومی و چهارمین کلمه‌ی پرتکرار مکالمه که با تقسیم بر تعداد کل کلمه‌های مکالمه نرمال شده‌اند.

ویژگی‌های ۱۲ تا ۱۴: به ترتیب سه کلمه پرتکرار مکالمه که با تقسیم بر تعداد کلمه‌های منحصر به فرد نرمال شده‌اند.

ویژگی ۱۸: میانگین شباهت رابطه‌ی (۳-۵) بین دو نوبت متوالی در تمام مکالمه است.

ویژگی ۱۹: میانگین شباهت کسینوسی نوبت‌های پشت سر هم است.

ویژگی ۲۱: آنتروپی خوشه‌های ایجاد شده در خوشه‌بندی بر اساس معیار کسینوسی.

ویژگی ۲۳: آنتروپی خوشه‌ها در خوشه‌بندی با معیار رابطه (۵-۳).

۴-۳- ویژگی‌های ارتباط معنایی

همان‌طور که گفته شد به دلیل ارتباط معنایی بین کلمه‌ها و عدم تکرار کلمه‌های کلیدی، میزان انسجام مکالمه‌ها را تنها از طریق محاسبه تکرار کلمه‌های کلیدی به دست آورده نمی‌شوند. استفاده از یک پایگاه داده معنایی ممکن است باعث بهبود تشخیص میزان انسجام گردد.

روش‌های جانشانی کلمه مانند GLOVE [۲۹] با استفاده از شبکه عصبی، هر کلمه را به یک بردار نسبت می‌دهند. در این روش‌ها آموزش شبکه عصبی با داده‌های بسیار زیاد باعث می‌شود، مدل شبکه‌ی عصبی روابط پیچیده بین کلمات را یاد بگیرد. روابط بین کلمه‌ها با استفاده از بردارها نمایش داده می‌شوند. هر کلمه به یک بردار چند صدتایی نسب داده شده است. در GLOVE هرچه ارتباط کلمه‌ها بیشتر باشد، شباهت دو بردار هم بیشتر می‌شود. همچنین تفاوت دو بردار به عنوان مثال خواهر و برادر مشابه اختلاف دو بردار زن و مرد می‌باشد. در این پژوهش از پایگاه داده تولید شده توسط مدل GLOVE آموزش دیده شده روی crow common با B۸۴۰ tokens و ۲/۲ میلیون

کلمه برای محاسبه همنشینی کلمه‌ها استفاده شده است [۲۹]. این پایگاه داده شامل ۲/۲ میلیون کلمه می‌باشد. هر کلمه با یک بردار ۳۰۰ بعدی نمایش داده شده است.

شباهت بردار هر دو کلمه در این پایگاه داده، میزان شباهت و یا همنشین بودن آن‌ها می‌باشد. مانند مقاله‌ی [۲۴] از اندازه شباهت کسینوسی دو بردار استفاده شده است. هر چه اندازه شباهت بیشتر باشد میزان همنشین و یا ارتباط معنایی دو کلمه بیشتر می‌شود. مانند شباهت کسینوسی کلمه‌های "boy" و "girl" که برابر ۰/۸۱۴ می‌باشند.

۱-۳-۴- تعیین مقادیر ثابت

دو کلمه در صورتی همنشین گفته می‌شوند که شباهت کسینوسی دو بردار آن‌ها از مقدار مشخصی بیشتر باشد. ویژگی‌های مربوط به همنشینی کلمه‌ها، با حداقل مقدارهای متفاوتی از مجموعه آموزش استخراج می‌شوند. مقداری که بیشترین بهره اطاعات را در هر ویژگی داشته باشد به عنوان حداقل مقدار برای میزان همنشینی آن ویژگی انتخاب می‌شود.

جدول (۴-۶) مربوط به ویژگی شماره ۲۴ می‌باشد که با حداقل میانگین‌های همنشین‌های متفاوت از مجموعه آموزش استخراج شده است. ویژگی شماره ۲۴ میانگین تعداد نوبت‌هایی می‌باشد که با نوبت قبلی، میانگین همنشینی آن‌ها حداقل از مقدار θ بیشتر باشد. بیشترین بهره اطلاعات مربوط به همنشینی بیشتر از ۰/۷ می‌باشد.

جدول (۴-۶): بهره اطلاعات ویژگی شماره ۲۴ با حداقل همنشینی θ

θ	۰/۱	۰/۲	۰/۳	۰/۴	۰/۵	۰/۶	۰/۷	۰/۸	۰/۹
بهره اطلاعات	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰۱	۰/۰۳	۰/۰۲	۰/۰۲

جدول (۴-۷) ویژگی‌های شماره ۲۶ و ۲۷ را نشان می‌دهد. میانگین و آنتروپی خوشه‌ها در خوشه بندی بر اساس میانگین همنشینی نوبت‌ها با خوشه‌ها می‌باشد. تمام کلمه‌ها باهم مقداری همنشینی

دارند. برای آنکه تمام نوبت‌ها به یک خوشه تعلق پیدا نکنند و به‌طور صحیح خوشه‌بندی انجام شود، یک مقدار حداقل برای میانگین همنشینی تعریف شده است. با حداقل میانگین‌های متفاوت، خوشه-بندی‌های متفاوتی شکل گرفته است. بهره اطلاعات این خوشه‌بندی‌ها در جدول (۴-۷) نشان داده شده است. بهترین اندازه برای حداقل همنشینی در این ویژگی مقدار ۰/۳ می‌باشد.

جدول (۴-۷): بهره اطلاعات ویژگی‌های شماره ۲۶ و ۲۷ با حداقل همنشینی θ

θ		۰/۱	۰/۲	۰/۳	۰/۴	۰/۵	۰/۶	۰/۷	۰/۸	۰/۹
بهره اطلاعات	میانگین	۰/۰	۰/۰۷	۰/۱	۰/۰۸	۰/۰۵	۰/۰۳	۰/۰۱	۰/۰۶	۰/۰
	آنتروپی	۰/۰	۰/۰۷	۰/۰۹	۰/۰۶	۰/۰۴	۰/۰۳	۰/۰۱	۰/۰	۰/۰

ویژگی‌های شماره ۲۹ و ۳۰ به ترتیب میانگین و واریانس تعداد کلمه‌های همنشین در هر نوبت با تاریخچه مکالمه می‌باشند. در جدول (۴-۸) بهره اطلاعات این دو ویژگی با حداقل همنشینی‌های متفاوت آمده است. همنشینی ۰/۹ بهترین بهره اطلاعات را دارد، پس حداقل مقدار همنشینی ۰/۹ می‌باشد.

جدول (۴-۸): بهره اطلاعات ویژگی‌های شماره ۲۹ و ۳۰ با حداقل همنشینی θ

θ		۰/۱	۰/۲	۰/۳	۰/۴	۰/۵	۰/۶	۰/۷	۰/۸	۰/۹
بهره اطلاعات	میانگین	۰/۰۳	۰/۰۵	۰/۰۶	۰/۰۵	۰/۰۵	۰/۰۵	۰/۱۲	۰/۲۴	۰/۲۷
	واریانس	۰/۰۲	۰/۰۱	۰/۰۱	۰/۰۱	۰/۰۱	۰/۰۱	۰/۰۲	۰/۱۴	۰/۱۸

ویژگی ۳۱ میانگین تعداد نوبت‌هایی است که با حداقل ۱۰ درصد از کلمه‌های تاریخچه همنشین باشند. بهره اطلاعات این ویژگی با حداقل مقدار همنشینی‌های متفاوت در جدول (۴-۹) آمده است. بیشترین بهره اطلاعات مربوط به حداقل همنشینی ۰/۹ می‌باشد.

جدول (۴-۹): بهره اطلاعات ویژگی شماره ۳۱ با حداقل همنشینی θ

θ	۰/۱	۰/۲	۰/۳	۰/۴	۰/۵	۰/۶	۰/۷	۰/۸	۰/۹
بهره اطلاعات	۰/۰۱	۰/۰	۰/۰۱	۰/۰۳	۰/۰۵	۰/۰۸	۰/۱۳	۰/۲۴	۰/۲۷

ویژگی ۳۲ میانگین تعداد نوبت‌هایی است که با حداقل ۵ و حداکثر ۱۰ درصد از کلمه‌های تاریخچه همنشین باشند. بهره اطلاعات این ویژگی با حداقل همنشینی‌های متفاوت در جدول (۴-۱۰) آمده است. بهترین مقدار برای حداقل همنشینی ۰/۹ می‌باشد.

جدول (۴-۱۰): بهره اطلاعات ویژگی شماره ۳۲ با حداقل همنشینی θ

θ	۰/۱	۰/۲	۰/۳	۰/۴	۰/۵	۰/۶	۰/۷	۰/۸	۰/۹
بهره اطلاعات	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰۴	۰/۱۱

ویژگی ۳۳ میانگین تعداد نوبت‌هایی است که حداقل ۴۰ درصد از کلمه‌های موجود در آن‌ها با کلمه-های تاریخچه‌ی گفتگو همنشین باشند. بهره اطلاعات این ویژگی نیز در جدول (۴-۱۱) آمده است. مشابه با ویژگی‌های ۳۱ و ۳۲ بهترین مقدار برای حداقل همنشینی ۰/۹ است.

جدول (۴-۱۱): بهره اطلاعات ویژگی شماره ۳۳ با حداقل همنشینی θ

θ	۰/۱	۰/۲	۰/۳	۰/۴	۰/۵	۰/۶	۰/۷	۰/۸	۰/۹
بهره اطلاعات	۰/۰	۰/۰	۰/۰۱	۰/۰۲	۰/۰۵	۰/۰۶	۰/۱	۰/۱۹	۰/۲۱

سه ویژگی انتهایی که در فصل قبل معرفی شد مربوط به دسته‌بندی مکالمه می‌باشد. مکالمه‌های غیر منسجم از تغییر در نوبت‌های مکالمه به وجود می‌آیند. پس می‌توان با شناخت نوبت‌های شکسته شده به میزان انسجام مکالمه پی برد. نحوه شناسایی و ویژگی‌های استفاده شده برای شناخت شکست‌ها در فصل قبل توضیح داده شده است. از الگوریتم درخت تصمیم برای دسته‌بندی نوبت‌های مکالمه به دو دسته‌ی شکست و اتصال استفاده می‌شود. ویژگی‌های استفاده شده مربوط به همنشینی کلمه‌ها می‌باشند. حداقل مقدار همنشینی دو کلمه باید به‌گونه‌ای باشد که باعث بهتر شدن دسته‌بندی نوبت‌ها شود. در جدول (۴-۱۲) میانگین معیار اندازه‌گیری f دسته‌بندی با الگوریتم درخت تصمیم روی قسمتی از مجموعه آموزش با حداقل همنشینی‌های متفاوت نشان داده شده است. این مجموعه در آموزش الگوریتم دسته‌بندی نمی‌باشد. در این دسته‌بندی حداقل همنشینی که بهترین نتیجه دسته‌بندی را داشته است، ۰/۹بوده است.

جدول (۴-۱۲): نتیجه دسته‌بندی نوبت‌های مکالمه با حداقل همنشینی θ

θ	۰/۵	۰/۶	۰/۷	۰/۸	۰/۹
معیار اندازه‌گیری f	۰/۶۲۶	۰/۶۴۷	۰/۶۵	۰/۶۶۷	۰/۶۹۹

بهره اطلاعات ۱۲ ویژگی مربوط به همنشینی کلمه‌ها در جدول (۴-۱۳) آمده است. بعضی از ویژگی‌ها، بهره اطلاعات کمی دارند. این ویژگی‌ها در دسته‌بندی تأثیر زیادی ندارند. به همین دلیل آن‌ها را حذف می‌کنیم. ویژگی‌های شماره ۲۶، ۲۷، ۲۹، ۳۰، ۳۲ و ۳۶ حذف و ویژگی‌های شماره ۲۴، ۲۵، ۲۸، ۳۱، ۳۳، ۳۴ و ۳۵ انتخاب می‌شوند.

جدول (۴-۱۳): بهره اطلاعات ویژگی‌ها

شماره ویژگی	بهره اطلاعات
۲۴	۰/۲۸
۲۵	۰/۲۷
۲۶	۰/۰۳
۲۷	۰/۰۶
۲۸	۰/۲۷
۲۹	۰/۱۰
۳۰	۰/۰۹
۳۱	۰/۲۷
۳۲	۰/۱۹
۳۳	۰/۲۱
۳۴	۰/۲۷
۳۵	۰/۲۵
۳۶	۰/۱۱

۴-۳- همبستگی اندازه‌ی انسجام و ویژگی‌ها

تعداد ۲۱ ویژگی انتخاب شده برای دسته‌بندی مکالمه‌ها به شرح زیر می‌باشند:

ویژگی ۱: میانگین تعداد خوشه‌ها در خوشه‌بندی بر اساس بیشترین تعداد تکرار کلمه‌ها.

ویژگی ۲: آنتروپی خوشه‌ها در خوشه‌بندی بر اساس بیشترین تعداد تکرار کلمه‌ها.

ویژگی ۴: میانگین تعداد نوبت‌هایی که با حداقل ۵ و حداکثر ۱۰ درصد از کلمه‌های موجود در تاریخچه گفتگو تکراری باشند.

ویژگی ۶: میانگین تعداد نوبت‌هایی که حداقل یک کلمه تکراری با تاریخچه گفت گو دارند.

ویژگی ۷: واریانس تعداد نوبت‌هایی که حداقل یک کلمه تکراری با تاریخچه گفت گو دارند.

ویژگی‌های ۹ و ۱۱: به ترتیب دومی و چهارمین کلمه‌ی پرتکرار مکالمه که با تقسیم بر تعداد کل کلمه‌های مکالمه نرمال شده‌اند.

ویژگی‌های ۱۲ تا ۱۴: به ترتیب سه کلمه پرتکرار مکالمه که با تقسیم بر تعداد کلمه‌های منحصر به فرد نرمال شده‌اند.

ویژگی ۱۸: میانگین شباهت رابطه‌ی (۳-۵) بین دو نوبت متوالی در تمام مکالمه است.

ویژگی ۱۹: میانگین شباهت کسینوسی نوبت‌های پشت سر هم است.

ویژگی ۲۱: آنتروپی خوشه‌های ایجاد شده در خوشه‌بندی بر اساس معیار کسینوسی.

ویژگی ۲۳: آنتروپی خوشه‌ها در خوشه‌بندی با معیار رابطه (۵-۳).

ویژگی ۲۴: میانگین تعداد نوبت‌هایی که با نوبت بعدی، میانگین همنشینی بین کلمه‌هایشان بیشتر از ۰/۷ است.

ویژگی ۲۵: میانگین، میانگین همنشینی هر نوبت با تمام نوبت‌های گذشته.

ویژگی ۲۸: میانگین، میانگین همنشینی کلمه‌ها بین دو نوبت متوالی می‌باشد.

ویژگی ۳۱: میانگین تعداد نوبت‌هایی که با حداقل ۱۰ درصد از کلمه‌های موجود در تاریخچه گفتگو همنشینی بیشتر از ۰/۹ داشته باشند.

ویژگی ۳۳: میانگین تعداد نوبت‌هایی که حداقل ۴۰ درصد از کلمه‌هایشان با کلمه‌های موجود در تاریخچه گفتگو همنشینی بیشتر از ۰/۹ داشته باشند.

ویژگی ۳۴: میانگین تعداد نوبت‌هایی است که در دسته‌بندی مکالمه، شکست پیش‌بینی شده‌اند.

ویژگی ۳۵: آنتروپی دسته‌ها، در دسته‌بندی نوبت‌های مکالمه.

برای اثبات صحت این ۲۱ ویژگی و رابطه آن‌ها با اندازه‌ی انسجام، می‌توان با مقدار همبستگی، آن‌ها را نشان داد. همان‌طور که گفته شد مکالمه‌های غیر منسجم با ترکیب مکالمه‌ها و جانشین کردن نوبت‌ها ساخته شده است. فرض کنید تعداد نوبت‌های صحیح، تقسیم بر تعداد کل نوبت‌های مکالمه، اندازه انسجام یک مکالمه باشد. همبستگی زیاد بین مقادیر ویژگی‌ها و اندازه انسجام نشان از صحیح بودن ویژگی‌ها دارد. همبستگی، حاصل تقسیم کواریانس بر ضرب انحراف مقادیر دو ویژگی می‌باشد. در رابطه (۲-۴) همبستگی بین دو داده آمده است که میزان وابسته بودن دو داده را نسبت به هم مشخص می‌کند.

$$corr(X.Y) = \frac{cov(X.Y)}{\sigma_x \sigma_y} \qquad (۲-۴)$$

$$cov(X.Y) = E\left[(X - \mu_x)(Y - \mu_y)\right] \qquad (۳-۴)$$

با مقادیر مجموعه‌ی آموزش، همبستگی بین ویژگی‌ها و اندازه انسجام به دست می‌آید که نتایج آن در جدول (۴-۱۴) نشان داده شده است. میزان همبستگی منفی به معنای رابطه عکس بین ویژگی و اندازه انسجام است. افزایش مقدار ویژگی باعث کاهش اندازه انسجام می‌شود، مانند ویژگی شماره ۱ که هرچه تعداد نوبت‌های هر خوشه بیشتر باشد در واقع میانگین خوشه‌ها کمتر و انسجام بیشتر می‌باشد.

جدول (۴-۱۴): میزان همبستگی بین اندازه انسجام و ویژگی‌ها

همبستگی	شماره ویژگی
-۰/۵۸	۱
-۰/۶۲	۲
-۰/۵۴	۴
۰/۵۱	۶
۰/۴۱	۷
۰/۵۹	۹
۰/۴۹	۱۱
۰/۵۰	۱۲
۰/۵۲	۱۳
۰/۵۲	۱۴
۰/۴۹	۱۸
۰/۵۳	۱۹
-۰/۵۱	۲۱
-۰/۵۹	۲۳
-۰/۵۴	۲۴
-۰/۵۸	۲۵
۰/۵۷	۲۸
۰/۳۷	۳۱
۰/۳۶	۳۳
۰/۵۸	۳۴
۰/۵۲	۳۵

۴-۳- دسته‌بندی مکالمه

در این بخش ویژگی‌های پیشنهاد شده برای دسته‌بندی مکالمه در سه سطح عالی، متوسط و ضعیف مورد ارزیابی قرار می‌گیرند. از چند الگوریتم برای دسته‌بندی مکالمه‌ها استفاده شده است تا میزان

کارایی مجموعه ویژگی‌های پیشنهاد شده نشان داده شود. روش‌های دسته‌بندی مورد استفاده در این پژوهش بیزین ساده، ماشین بردار پشتیبان و درخت تصمیم می‌باشند. هر کدام از روش‌ها با استفاده از نرم افزار وکا و داده‌های مجموعه‌ی آموزش، آموزش داده‌شده‌اند. سپس با مجموعه‌ی آزمون، میزان کارایی ویژگی‌ها و صحت الگوریتم‌های دسته‌بندی مورد بررسی قرار گرفته‌اند.

۴-۳-۱- نتایج حاصل از دسته‌بندی

هر یک از روش‌های دسته‌بندی ذکر شده با مجموعه آموزش، آموزش می‌بیند که نتایج مجموعه آزمون هر یک از آن‌ها در جدول (۴-۱۵) آمده است. از میانگین معیار اندازه‌گیری f برای مقایسه و نمایش نتایج حاصل از دسته‌بندی استفاده شده است. در واقع معیار اندازه‌گیری f یک نوع میانگین بین صحت[1] و بازیابی[2] دسته‌بندی است. صحت سیستم، تعداد داده‌های صحیح پیش‌بینی شده به‌کل داده‌های پیش‌بینی شده است. بازیابی، نسبت تعداد داده های صحیح پیش‌بینی شده به‌کل داده‌های مورد انتظار برای پیش‌بینی می‌باشد.

هر یک از الگوریتم‌های دسته‌بندی با ۳ مجموعه از ویژگی‌ها آموزش می‌بیند و نتایج آن‌ها مورد مقایسه قرار می‌گیرد. مجموعه‌ی اول ویژگی‌های مربوط به تکرار کلمه‌ها، مجموعه‌ی دوم ویژگی‌های منتخب مجموعه‌ی تکرار کلمه‌ها و مجموعه‌ی سوم ویژگی‌های منتخب مربوط به تکرار کلمه‌ها و منتخب مربوط به همنشینی کلمه‌ها می‌باشد که مورد ارزیابی قرار می‌گیرند. شماره ویژگی‌های هر ۳ مجموعه در زیر آمده است.

مجموعه اول: ۱ تا ۲۴

مجموعه دوم: ۱، ۲، ۴، ۶، ۷، ۹، ۱۱، ۱۲، ۱۳، ۱۴، ۱۸، ۱۹، ۲۱ و ۲۳

[1] Perecision
[2] Recall

مجموعه سوم: ۱، ۲، ۴، ۶، ۷، ۹، ۱۱، ۱۲، ۱۳، ۱۴، ۱۸، ۱۹، ۲۱، ۲۳، ۲۴، ۲۵، ۲۸، ۳۱، ۳۳، ۳۴ و ۳۵

جدول (۴-۱۵) : نتایج دسته‌بندی مجموعه‌های آزمون

	بیزین	درخت تصمیم	SVM
مجموعه اول	۶۱/۸٪	۵۸٪	۶۱/۸٪
مجموعه دوم	۶۲/۸٪	۶۰/۹٪	۶۲/۸٪
مجموعه سوم	۶۳٪	۶۱/۴٪	۶۴٪

جدول (۴-۱۵) نشان دهنده بهبود دسته‌بندی با استفاده از انتخاب ویژگی (مجموعه‌ی دوم) و ویژگی-های مربوط به همنشینی (مجموعه‌ی سوم) می‌باشد.

در هر سه الگوریتم استفاده شده، مجموعه‌ی دوم نسبت به مجموعه‌ی اول نتیجه بهتری در دسته-بندی داشته است. ویژگی‌های مربوط به تکرار کلمه‌های کلیدی در حالت‌های مختلف از مکالمه‌ها استخراج شده است، مانند خوشه‌بندی مکالمه که با سه معیار متفاوت خوشه‌بندی می‌شود. این ویژگی‌ها ممکن است مقادیر مشابه داشته باشند، در واقع یک ویژگی تکرار شده باشد. مقادیر تکرار شده تأثیری در الگوریتم دسته‌بندی ندارد و ویژگی اضافه می‌باشند. با شناسایی و حذف این ویژگی‌ها زمان اجرای الگوریتم سریع‌تر می‌شود.

مقادیر بعضی دیگر از ویژگی‌ها در تمام مکالمه‌ها با سطوح مختلف به‌گونه‌ای است که بین دسته‌ها تفاوتی نمی‌باشند. در واقع مقادیر مشابه در دسته‌های مختلف زیاد است. این نوع ویژگی‌ها باعث کاهش کارایی دسته‌بندی می‌شود. به دو علت مقادیر این ویژگی‌ها به وجود آمده‌اند. برای ساخت مکالمه‌های غیر منسجم از ترکیب خودکار مکالمه‌ها استفاده شده است. این روش باعث شده است که مکالمه‌ها با موضوعات مشابه باهم ترکیب شوند. علت دیگر، مکالمه‌های منسجم می‌باشند که کلمه-

های کلیدی کمی در آن‌ها تکرار شده است. تشخیص انسجام این نوع مکالمه‌ها فقط با استفاده از همنشینی کلمه‌ها به دست می‌آید.

به دلیل آنکه الگوریتم SVM در هر سه مجموعه نسبت به دو الگوریتم دیگر بهترین نتیجه‌ی دسته-بندی را داشته است، نتایج این دسته‌بندی را بررسی می‌کنیم. با استفاده از مجموعه همنشینی کلمه-ها، همان‌طور که انتظار می‌رفت میانگین تشخیص صحیح مکالمه‌ها افزایش پیدا کرد. تفاوت نتایج آزمون دسته‌بندی با مجموعه‌ی دوم و سوم در جدول (۴-۱۶) نشان داده شده است.

جدول (۴-۱۶): مقایسه نتایج پیش‌بینی الگوریتم دسته‌بند روی مجموعه‌های دوم و سوم

انسجام مکالمه	پیش‌بینی مجموعه دوم	پیش‌بینی مجموعه سوم	تعداد تغییرات
سطح A	B	A	۷
	A	B	۳
	B	C	۱
سطح B	A	B	۱
	B	C	۲
سطح C	C	B	۲
	B	C	۱

تعداد ۷ مکالمه که انسجام عالی داشته‌اند در مجموعه‌ی دوم، انسجام متوسط تشخیص داده شده‌اند. استفاده از همنشینی کلمه‌های کلیدی باعث شده است که این ۷ مکالمه با مجموعه‌ی سوم درست تشخیص داده شوند. تمامی این ۷ مکالمه، دارای نوبت‌های طولانی می‌باشند. مکالمه‌ها برحسب طول نوبت‌هایشان دارای تعداد کلمه‌های مشابه کمی می‌باشند. یکی از این هفت مکالمه، نمونه‌ی شماره ۲۰۹۰ می‌باشد که در شکل (۴-۱) آمده است. در این مکالمه تعداد کلمه‌های تکرار شده در نوبت‌ها

نسبت به طول نوبت‌ها کم می‌باشد. به همین دلیل فقط با استفاده از ویژگی‌های تکرار کلمه‌ها، این مکالمه به اشتباه سطح متوسط تشخیص داده می‌شود.

personA: Can you drive me to the airport tomorrow afternoon?

personB: I would be happy to drive you to the airport. When do you need to leave?

personA: For an international flight, the airline wants us to check in two hours in advance, so I need to leave my house at 2:00.

personB: Do you have a lot of luggage to check in?

personA: I am only carrying one large bag and a carry-on.

personB: I'll bring our smaller car then. Do you know where you need to get dropped off?

personA: Yes, I am flying out on Singapore Air to Sydney, so I need to be at the International Terminal.

personB: Will I be able to come in with you to say goodbye?

personA: Actually, with airport regulations being so strict, you should probably just drop me off at the curb.

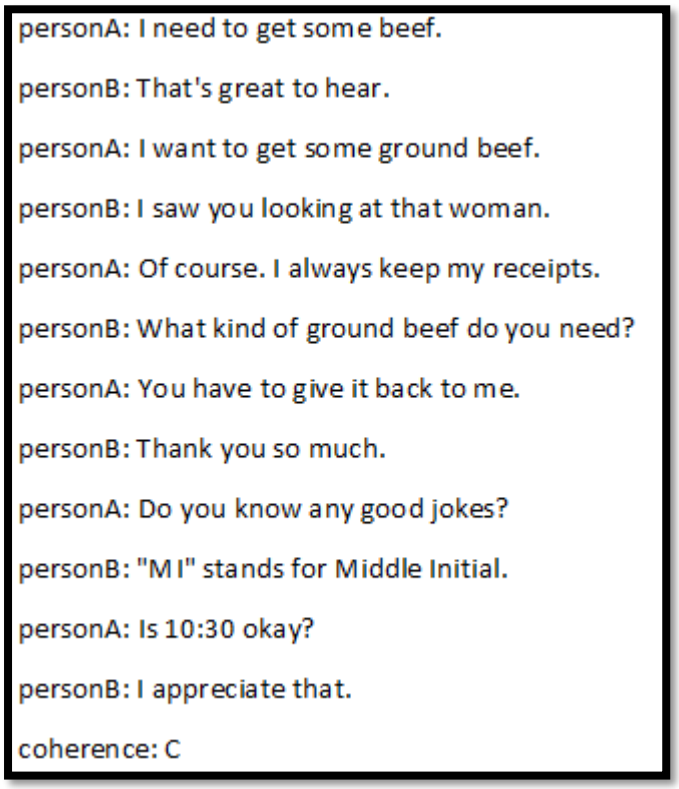
personB: That will work out fine. Just give me a call if anything changes.

coherence: A

شکل (۴-۱): مکالمه با انسجام عالی

در نتایج دیگر دسته‌ها تغییر زیادی به وجود نیامده، در نتیجه استفاده از مجموعه‌ی همنشینی در تشخیص مکالمه‌های سطح عالی مؤثر بوده است. انتظار می‌رفت ویژگی‌های مربوط به همنشینی بتواند دو دسته متوسط و ضعیف را بهتر از یکدیگر جدا کند، اما این‌گونه نبود. دلیل این امر می‌تواند مکالمه-های استفاده شده باشند. از مکالمه‌های آموزش زبان انگلیسی در این پژوهش استفاده شده است که این مکالمه‌ها محاوره‌ای و روزمره می‌باشند. در ساخت مکالمه با انسجام کمتر چند مکالمه‌ی روزمره باهم ترکیب می‌شوند. انتخاب مکالمه و نوبت‌ها برای ترکیب کاملا تصادفی و خودکار است. ممکن است دو مکالمه با موضوع‌های مشابه ترکیب شوند. به صورت کلی چون مکالمه‌ها محاوره‌ای می‌باشند پس ممکن است کلمه‌های نوبت جانشین، ارتباط معنایی با دیگر کلمه‌های مکالمه داشته باشد. در واقع کلمه‌های زیادی از نظر معنا به هم مرتبط و یا یکسان هستند که در مکالمه‌های روزمره با موضوع‌های مختلف حضور دارند. حضور این نوع کلمه‌ها در مکالمه‌های پایگاه داده، تشخیص انسجام را با استفاده از پایگاه داده معنایی مشکل می‌سازد. مانند نمونه مکالمه‌ی شماره‌ی ۲۰۱۹(شکل (۴-۲)) که با تغییر در نوبت‌های ۲، ۴، ۵، ۷، ۸، ۹، ۱۰ و ۱۱ به انسجام ضعیف رسیده است. اما هر سه

نوع دسته‌بندی با هر سه مجموعه این مکالمه را به اشتباه انسجام متوسط تشخیص داده‌اند. در این مکالمه تعداد کلمه‌های کلیدی در مکالمه‌ی اصلی (مکالمه‌ی بدون تغییر) زیاد تکرار شده است. با تغییر در نوبت‌ها، بازهم کلمه‌ی تکرار شده زیاد می‌باشد و انسجام متوسط تشخیص داده می‌شود.


personA: I need to get some beef.

personB: That's great to hear.

personA: I want to get some ground beef.

personB: I saw you looking at that woman.

personA: Of course. I always keep my receipts.

personB: What kind of ground beef do you need?

personA: You have to give it back to me.

personB: Thank you so much.

personA: Do you know any good jokes?

personB: "MI" stands for Middle Initial.

personA: Is 10:30 okay?

personB: I appreciate that.

coherence: C

شکل (۴-۲): نمونه مکالمه‌ی شماره ۲۰۱۹

۴-۵- جمع‌بندی

در این فصل ویژگی‌های معرفی شده مورد ارزیابی قرار گرفته‌اند. ابتدا با استفاده از مجموعه‌ی آموزش و الگوریتم‌های انتخاب ویژگی، ویژگی‌هایی که مقادیر ثابت دارند به‌دست می‌آیند. سپس از بین تمام ویژگی‌های معرفی شده، زیر مجموعه‌ای از بهترین ویژگی‌ها برای ارزیابی انتخاب می‌شوند. ویژگی‌های استخراج شده از مکالمه‌ها با انواع الگوریتم‌های دسته‌بندی، به دو دسته منسجم و غیر منسجم تقسیم شده و میزان دقت این الگوریتم‌ها نشان از کارایی ویژگی‌ها در تعیین انسجام مکالمه دارند. سپس به تحلیل نتایج الگوریتم‌ها و تأثیر استفاده از همنشینی کلمه‌ها در تشخیص انسجام پرداخته شده است. بهترین نتیجه‌ی دسته‌بندی مربوط به دسته‌بندی با الگوریتم SVM روی ویژگی‌های تکرار کلمات

کلیدی و میزان همنشینی آن‌ها می‌باشد. نتیجه این دسته‌بندی روی مجموعه آزمون، میانگین معیار اندازه‌گیری f برابر ٪۶۴ می‌باشد.

فصل پنجم:

جمع‌بندی و پیشنهادهایی برای ادامه کار

۱-۶- جمع‌بندی

مکالمه یک نوع ارتباط دو طرفه بین دو انسان و یا انسان و سیستم می‌باشد. مکالمه بین انسان و سیستم به عنوان سیستم‌های گفتگو و یا عامل‌های مکالمه‌ای شناخته می‌شوند. عامل‌های مکالمه‌ای سیستم‌هایی هستن که از طریق زبان طبیعی با انسان ارتباط برقرار می‌کنند. اهداف سیستم‌های گفتگو صرفا برای سرگرمی و یا انجام وظیفه خاصی می‌باشند. این سیستم‌ها برای جلب رضایت کاربر و انجام هدف سیستم در کمترین زمان، لازم است پرسش و پاسخ‌های تولید شده توسط سیستم‌ها مانند پرسش و پاسخ‌های تولید شده توسط انسان خبره باشد. پرسش و پاسخ‌های تولید شده توسط انسان خبره، با موضوع مکالمه مرتبط و منسجم است. به عنوان مثال در سیستم‌های وظیفه گرا مانند پرسش و پاسخ تعاملی هرچه تولیدات سیستم به موضوع مکالمه و پرسش و پاسخ‌های گذشته نزدیک‌تر باشد، سریع‌تر کاربر به پاسخ مورد نظر می‌رسد. در واقع هرچه انسجام بین تولیدات سیستم و پرسش و پاسخ‌های گذشته بیشتر باشد، سیستم بهتر عمل می‌کند. تاکنون در سیستم‌های معرفی شده در این حوزه ذکر شده انسجام رعایت می‌شود. اندازه‌گیری دقت آن‌ها در رعایت انسجام نظر انسان می‌باشد. روشی اتوماتیک برای اندازه‌گیری انسجام معرفی نشده است.

در این پژوهش با استفاده از نحوه و میزان تکرار و همنشینی کلمه‌های کلیدی، روشی اتوماتیک برای اندازه‌گیری انسجام مکالمه معرفی شده است. ویژگی‌های استخراج شده با استفاده از الگوریتم‌های دسته‌بندی به سه دسته انسجام عالی، متوسط و ضعیف دسته‌بندی شده‌اند.

به‌طورکلی این پژوهش شامل سه بخش زیر می‌باشد.

۱- در حالت‌های مختلف تعداد تکرار کلمه‌های کلیدی محاسبه می‌شود. ویژگی‌های استخراج شده، چون در حالت‌های مختلف محاسبه شده‌اند به هم وابسته می‌باشند.

۲- برای حذف ویژگی‌های نامناسب از الگوریتم انتخاب ویژگی استفاده شده. وجود ویژگی‌های اضافی و نویزی باعث کاهش کارایی الگوریتم‌های دسته‌بندی می‌شوند. با شناسایی ویژگی‌های نامناسب و حذف آن‌ها معیار اندازه‌گیری f تشخیص الگوریتم دسته‌بند افزایش پیدا کرده است.

۳- استفاده از میزان همنشینی کلمه‌ها در محاسبه ویژگی‌های جدید به منظور افزایش شناسایی مکالمه‌های منسجم می‌باشد. این ویژگی‌ها با ویژگی‌های قبلی ترکیب شده و باعث بهبود شناسایی مکالمه‌ها با انسجام عالی شده است.

۲-۶- کارهای آینده

در این پژوهش برای بررسی میزان همنشینی کلمه‌ها از پایگاه داده GLOVE موجود استفاده شده است. همنشینی‌های استفاده شده مربوط به متن‌ها با موضوعات مختلف می‌باشد. برای بهبود تشخیص میزان انسجام، می‌توان پایگاه داده همنشینی را از روی مکالمه‌ها ساخته شود زیرا میزان همنشینی کلمه‌ها در مکالمه‌های محاوره‌ای با متن متفاوت می‌باشد.

همچنین برای بهبود تشخیص انسجام، می‌توان از مکالمه‌ها با موضوعات یکسان برای به عنوان پایگاه داده برای آموزش استفاده کرد.

منابع و مأخذ:

[1] A. Chopra, A. Prashar, and C. Sain, (2013), "Natural Language Processing," *International Journal of Technology Enhancements and Emerging Engineering Research*, pp. 131-134.

[2] M. Abdolahi and M. Zahedi, (2016), "An overview on text coherence methods," in *Information and Knowledge Technology (IKT), 2016 Eighth International Conference on*, pp. 1-5.

[3] S. A. Crossley, K. Kyle, and D. S. McNamara, (2016), "The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion," *Behavior Research Methods*, vol. 48, pp. 1227-1237.

[4] W. Wei, Z. Ming, L. Nie, G. Li, J. Li, F. Zhu, *et al.*, (2016), "Exploring heterogeneous features for query-focused summarization of categorized community answers," *Information Sciences* ,vol. 330, pp. 403-423.

[5] J. Christensen, S. S. Mausam, S. Soderland, and O. Etzioni, (2013), "Towards Coherent Multi-Document Summarization," in *HLT-NAACL*, pp. 1163-1173.

[6] R. Zhang, W. Li, N. Liu, and D. Gao, (2016), "Coherent narrative summarization with a cognitive model," *Computer Speech & Language*, vol. 35, pp. 134-160.

[7] G. Leroy, D. Kauchak, and O. Mouradi, (2013), "A user-study measuring the effects of lexical simplification and coherence enhancement on perceived and actual text difficulty," *International Journal of Medical Informatics*, vol. 82, pp. 717-730.

[8] D. Xiong, Y. Ding, M. Zhang, and C. L. Tan, (2013), "Lexical chain based cohesion models for document-level statistical machine translation," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1563-1573.

[9] G. Carenini, R. T. Ng, and X. Zhou, (2008), "Summarizing Emails with Conversational Cohesion and Subjectivity," in *ACL*, pp. 353-361.

[10] T. Adams, (2017), "AI-Powered Social Bots," *arXiv preprint arXiv:1706.05143*.

[11] D. Jurafsky and J. H. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition," ed: MIT Press.

[12] K. H. Kendrick, (2017), "Using Conversation Analysis in the Lab," ed: Taylor & Francis.

[13] C. Chakrabarti, (2014), "*Artificial conversations for chatter bots using knowledge representation, learning, and pragmatics"* The University of New Mexico.

[14] A. Bartl and G. Spanakis, (2017), "A retrieval-based dialogue system utilizing utterance and context embeddings," *arXiv preprint arXiv:1710.05780*.

[15] K. Natalia and O. Constantin, (2013),"Interactive Question Answering," in *Emerging Applications of Natural Language Processing: Concepts and New Research*, B. Sivaji, N. Sudip Kumar, and E. Asif, Eds., ed Hershey, PA, USA: IGI Global, pp. 149-169.

[16] B. Sacaleanu, C. Orasan, C. Spurk, S. Ou, O. Ferrandez, M. Kouylekov, *et al.*, (2008), "Entailment-based question answering for structured data," in *22nd International Conference on on Computational Linguistics: Demonstration Papers*, pp. 173-176.

[17] R. Bernardi and M. Kirschner, (2008), "Context modeling for iqa: the role of tasks and entities,", in *Coling 2008: Proceedings of the workshop on Knowledge and Reasoning for Answering Questions*, pp. 25-32.

[18] T. Mori, S. Kawaguchi, and M. Ishioroshi, (2007), "Answering contextual questions based on the cohesion with knowledge," *International Journal of Computer Processing of Oriental Languages*, vol. 20, pp. 115-135.

[19] H. Liu, T. Lin, H. Sun, W. Lin, C.-W. Chang, T. Zhong, *et al.*, (2017), "RubyStar: A Non-Task-Oriented Mixture Model Dialog System," *arXiv preprint arXiv:1711.02781*.

[20] R. Barzilay and M. Lapata, (2005), "Modeling local coherence: an entity-based approach," presented at the Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, Ann Arbor, Michigan.

[21] R. Barzilay and M. Lapata, (2008), "Modeling local coherence: An entity-based approach," *Computational Linguistics*, vol. 34, pp. 1-34.

[22] J. Burstein, J. Tetreault, and S. Andreyev, (2010), "Using entity-based features to model coherence in student essays," in *Human language technologies: The 2010 annual conference of the North American chapter of the Association for Computational Linguistics*, pp. 681-684.

[23] C. Guinaudeau and M. Strube, (2013), "Graph-based Local Coherence Modeling," in *ACL (1)*, pp. 93-103.

[24] M. Mesgar and M. Strube, (2016), "Lexical Coherence Graph Modeling Using word embeddings," in *HLT-NAACL*, pp. 1414-1423.

[25] M. Flor and B. B. Klebanov, (2014), "Associative lexical cohesion as a factor in text complexity," *ITL-International Journal of Applied Linguistics*, vol. 165, pp. 223-258.

[26] P. W. Foltz, W. Kintsch, and T. K. Landauer, (1998), "The measurement of textual coherence with latent semantic analysis," *Discourse processes*, vol. 25, pp. 285-307.

[27] S. A. Crossley, P. M. McCarthy, T. Salsbury, and D. S. McNamara, (2008), "LSA as a measure of coherence in second language natural discourse," in *Proceedings of the 30th annual conference of the Cognitive Science Society*, pp. 1906-1911.

[28] G. Carenini, R. T. Ng, and X. Zhou, (2007), "Summarizing email conversations with clue words," in *Proceedings of the 16th international conference on World Wide Web*, pp. 91-100.

[29] J. Pennington, R. Socher, and C. Manning, (2014), "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532-1543.

[30] M. Mesgar and M. Strube, (2015), "Graph-based Coherence Modeling For Assessing Readability," in *NAACL-HLT*, pp. 309-318.

[31] P. Modaresi, M. Liebeck, and S. Conrad, (2016), "Neural Classification of Linguistic Coherence using Long Short-Term Memories," in *Proceedings of the 8th annual meeting of the Forum on Information Retrieval Evaluation*, pp. 28-31.

[32] J. Li and D. Jurafsky, (2016), "Neural net models for open-domain discourse coherence," *arXiv preprint arXiv:1606.0154*.

[33] J. Li and E. H. Hovy, (2014), "A Model of Coherence Based on Distributed Sentence Representation," in *EMNLP*, pp. 2039-2048.

[34] H. Mei, M. Bansal, and M. R. Walter, (2017), "Coherent Dialogue with Attention-Based Language Models," in *AAAI*, pp. 3252-3258.
[35] S. Deng, P. Zhang, and Y. Zhou, (2011), "Turning unstructured and incoherent group discussion into DATree: A TBL coherence analysis approach".
[36] K. W. Church and P. Hanks, (1990), "Word association norms, mutual information, and lexicography," *Computational linguistics*, vol. 16, pp. 22-29.
[37] G. Bouma, (2009), "Normalized (pointwise) mutual information in collocation extraction," *Proceedings of GSCL*, pp. 31-40.
[38] A. Fang, C. Macdonald, I. Ounis, and P. Habel, (2016), "Using word embedding to evaluate the coherence of topics from Twitter data," in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 1057-1060.
[39] D. Xiong and M. Zhang, (2013), "A Topic-Based Coherence Model for Statistical Machine Translation," in *AAAI*.
[40] M. Lapata and R. Barzilay, (2005), "Automatic evaluation of text coherence: Models and representations," in *IJCAI*, pp. 1085-1090.
[41] M. A. Hall, (1999), "Correlation-based feature selection for machine learning".

Abstract

Conversation is a type of two-way communication between two humans. From the perspective of psychologists, a correct conversation has features that one of them is coherence. Coherence in conversation means keeping path, lack of gaps and sudden changes in the subject of the conversation. Changing the subject in a coherent conversation happens as fluent and relevant to past subjects.

A conversation can be connected between humans and system, such as dialogue and interactive question answering systems.

In order to increase the efficiency and popularity of the dialogue systems, the design of these systems should be such that the correct conversational criteria between the two humans are met. In fact, the dialogue systems must be respond, like an expert human. The correct conversation between two humans has the characteristics that their observance in dialogue systems, can increase coherence. Features, are identifiable by comparing coherent and non-coherent conversations. By recognizing the characteristic in coherent and non-coherent conversations, can be measure the coherence of a conversation.

So far, in built-conversational systems, coherence have been observed on them. The designers of these systems are used human opinions for to check the coherence of conversations. Use of human opinion is useful, but human cannot give the score to the systems, with the same opinion. So far, the automated system has not been designed, to be able give scores to the conversations with a same opinion.

In this research, by investigating the coherent and non-coherent produced conversations by human, the features, have been suggested to assess the level of coherence of a conversation. The features are divided into two groups of repetition and collocation of keywords. Each category of introduced features has been tested on the conversation series. The F measure derived from the conversation categorization with related features to key words repetition with support vector machine is 62.8%. The combination of the repetition and the collocation of the key words, has led to the F measure to 64%.

Keywords: coherence in conversation, repetition the key words in a coherent conversation, the effect of words collocation in coherence, the production of a conversation without coherence

Shahrood University of Technology

Faculty of Computer Engineering

M.Sc. Thesis in Artificial Intelligence Engineering

Coherence evaluation of a conversation using artificial intelligence techniques

By:
Fateme Zakizade gharie ali

Supervisor:
Dr. Morteza Zahedi

July 2018