



دانشگاه جامع علمی-کاربردی
مرکز آموزش عالی علمی-کاربردی جهاد دانشگاهی مشهد

اطلاعات فناوری مهندسی رشته کارشناسی نامه پایان
گرایش برنامه نویسی تحت وب

انسجام متن با استفاده از الگوریتم های یادگیری عمیق

استاد راهنما: جناب آقای دکتر عبدالهی

استاد مشاور: جناب آقای مهندس اخلاقی

نگارنده: فاطمه زرینی

زمستان 96

تشکر و قدردانی

از دوست و همراه گرامی جناب آقای بهزاد علیشاهی که در این مدت مرا یاری کردند کمال تشکر و قدردانی را دارم و از خداوند علو درجات را برای ایشان خواستارم. همچنین از استاد محترم آقای دکتر عبدالهی که مشوق بنده بودند و مسیری جدید در علوم کامپیوتر را به بنده نشان دادند کمال تشکر دارم.

چکیده

در سالیان اخیر توجه زیادی به روش‌های یادگیری عمیق شده است و موضوعات و مطالب زیادی در این زمینه وجود دارد که قابل پوشش دادن تحت یک عنوان کلی و در قالب این پایان نامه نمی‌باشد. یادگیری عمیق شاخه‌ای از بحث یادگیری ماشینی و مجموعه‌ای از الگوریتم‌هایی است که تلاش می‌کنند مفاهیم انتزاعی سطح بالا را با استفاده از یادگیری در سطوح و لایه‌های مختلف مدل کنند. یادگیری عمیق در واقع نگرشی جدید به ایده‌های شبکه‌های عصبی می‌باشد که سالیان زیادی است وجود داشته و هر چند سال یکبار در قالبی جدید خود را نشان می‌دهد. مگر انسان، به اذعان بسیاری از دانشمندان، پیچیده‌ترین سیستمی است که تا کنون در کل گیتی مشاهده شده و مورد مطالعه قرار گرفته است. اما این پیچیده‌ترین سیستم، نه ابعادی در حد کَشکشان دارد و نه تعداد اجزای سازنده اش، بیشتر از پردازنده‌های ابرکامپیوترهای امروزی است. پیچیدگی راز آلود این سیستم بی‌نظیر، به ارتباط‌های فراوان موجود میان عناصر آن بر می‌گردد. چیزی که، مغز ۱۴۰۰ گرمی انسان را، از همه سیستم‌های دیگر، متمایز می‌کند. با گسترش روز افزون متونی که توسط بشر روی بستر اینترنت منتشر می‌شود امروز بیش از پیش نیاز به روش‌های تحلیل و استخراج دانش و اطلاعات مفید احساس می‌شود و تقریباً تمامی شرکت‌ها و کمپانی‌ها روی به استفاده از روش‌های تحلیل زبانی خودکار آورده‌اند به همین دلیل اکنون روش‌های بسیاری برای تحلیل زبان‌های طبیعی وجود دارد و پیشرفت‌های بسیاری در زمینه‌ی حل مسائل آن به وجود آمده است در این پایان نامه می‌خواهیم از یادگیری عمیق در زمینه‌ی انسجام زبان‌های طبیعی استفاده کنیم.

کلمات کلیدی: یادگیری عمیق، انسجام متن، زبان طبیعی، پردازش متن

فهرست مطالب

چکیده..... 1

فصل اول

- 1-1- مقدمه 2
- 1-2- اهمیت و ضرورت پایان نامه 3
- 1-3- اهداف مشخص پایان نامه..... 3
- 1-4- سئوالات تحقیق..... 3
- 1-5- تعریف واژه ها و اصطلاحات فنی و تخصصی..... 4
- 1-5-1 یادگیری عمیق..... 4
- 1-5-2 شبکه های عصبی مصنوعی..... 4
- 1-5-3 شبکه های کانولوشن..... 4
- 1-5-4 بینایی ماشین..... 5
- 1-5-5 یادگیری ماشینی..... 5
- 1-5-6 شبکه های عصبی زیستی..... 6
- 1-5-7 پردازش زبان های طبیعی..... 6
- 1-5-8 پردازش و انسجام متن..... 6
- 1-5-9 پارامترهای مشخص کننده یک متن منسجم..... 7
- 1-5-10 یادگیری عمیق و بکارگیری آن در پردازش متن..... 10
- 1-5-11 انواع نگارش..... 11

فصل دوم

- 1-2- مقدمه..... 13
- 2-2- زبان‌های هوش مصنوعی 13
- 3-2- هوش مصنوعی AI 13
- 3-2- 1 تعریف‌های متفاوت برای هوش مصنوعی..... 13
- 2-3-2 ساختار مغز..... 14
- 4-2- یادگیری ماشین..... 15
- 1-4-2 مدل‌های یادگیری ماشین..... 16
- 1-5-2 یادگیری توسط شبکه‌های عصبی مصنوعی..... 17
- 2-5-2 ساختار شبکه‌های عصبی..... 19
- 3-5-2 کاربرد شبکه‌های عصبی..... 19
- 4-5-2 معایب شبکه‌های عصبی..... 20
- 5-5-2 در حالت کلی شبکه‌های عصبی به دو مفهوم کلی اشاره دارد..... 21
- 6-2- شبکه‌های عصبی زیستی..... 21
- 7-2- شبکه‌های عصبی مصنوعی..... 21
- 1-7-2 قابلیت‌های شبکه‌های عصبی..... 21
- 2-7-2 مزایای شبکه‌های عصبی..... 22
- 3-7-2 نورون مصنوعی..... 23
- 4-7-2 ساختار شبکه‌های عصبی..... 23
- 4-7-2 معایب شبکه‌های عصبی..... 25

فصل سوم

- 1-3- مقدمه..... 27
- 2-3- پیشینه تحقیق در یادگیری عمیق..... 27
- 3-3- یادگیری عمیق..... 28
- 4-3- شبکه‌های عصبی کانولوشن..... 30
- 5-3- انواع لایه‌های شبکه‌های کانولوشنی..... 31

32.....	1-5-3 لایه کانولوشن.....
33.....	2-5-3 لایه های pooling.....
34.....	3-5-3 لایه تماما متصل.....
35.....	6-3 معماری شبکه های عصبی کانولوشن.....
35.....	1-6-3 معماری AlexNet.....
35.....	2-6-3 معماری ZF.....
35.....	3-6-3 معماری GoogLeNet.....
36.....	4-6-3 معماری VGGNet.....
36.....	7-3 شبکه باور عمیق (DBN).....
37.....	8-3 شبکه عصبی feed forward.....
37.....	9-3 کاربردهای یادگیری عمیق.....
37.....	10-3 انسجام متن.....
38.....	1-10-3 رویکردهای مورد استفاده در تولید متن.....
38.....	2-10-3 رویکردهای مورد استفاده در خلاصه سازی متن.....
38.....	3-10-3 رویکردهای مورد استفاده در ساده سازی متن.....
39.....	4-10-3 رویکردهای مورد استفاده در امتیازدهی خودکار مقالات.....
39.....	5-10-3 رویکردهای مورد استفاده در ترجمه آماری ماشینی.....

41.....	منابع.....
---------	------------

فصل اول

کلیات

1-1 مقدمه

در سال 2006 میلادی روش یادگیری ساختار یافته عمیق یا به طور معمول یادگیری عمیق یا یادگیری سلسله مراتبی به عنوان یک شاخه جدید از علم بینایی ماشین پدیدار شد. در طی سال های گذشته، تکنیک های تعمیم یافته از روش یادگیری عمیق، تحت تاثیر طیف وسیعی از شاخه های علم پردازش اطلاعات و سیگنال مانند: هوش مصنوعی، یادگیری ماشین و همچنین پوشش رسانه ای بوده است. در سال های اخیر کارگاه ها، آموزش ها، نشریات یا جلسات خاص کنفرانس های متعددی به طور گسترده ای به زمینه یادگیری عمیق و کاربردهای آن در شاخه های مختلف علم پردازش اطلاعات و سیگنال پرداخته اند. این فرآیند در بسیاری از حالات شبکه های عصبی برای بهبود رفتارهایی از ماشین مانند تشخیص گفتار، بینایی کامپیوتر و پردازش زبان طبیعی استفاده می کند. یادگیری عمیق زیر مجموعه ای از یادگیری ماشین به حساب می آید که در آن بر اساس یادگیری چند سطحی در یک ساختار سلسله مراتبی از ویژگی ها یا مفاهیم سطوح بالاتر مفاهیم و ویژگی های سطوح پایین تر تعریف شده و همچنین مفاهیم سطح پایین نیز می توانند به تعریف مفاهیم سطح بالاتر کمک کنند. به زبان ساده تر هدف اصلی یادگیری عمیق استخراج ویژگی ها به صورت هوشمند طی چند مرحله یادگیری است. به طور خلاصه یادگیری عمیق را میتوان به این صورت معرفی کرد:

یادگیری عمیق شاخه ای از یادگیری ماشین و مجموعه ای از الگوریتم ها است که تلاش می کنند مفاهیم انتزاعی سطح بالا را با استفاده یادگیری در سطوح و لایه های مختلف مدل کند.

یادگیری عمیق نقطه مقابل مفهومی به نام یادگیری کم عمق¹ است. یادگیری کم عمق همان روش های معروف یادگیری ماشین است که در اغلب روش های پردازشی استفاده می شوند.

یادگیری عمیق اغلب از شبکه های عصبی برای بهبود رفتارهای ماشین مانند تشخیص گفتار²، بینایی کامپیوتر³ و پردازش زبان طبیعی⁴ استفاده می کند.

اولین مراحل معرفی و بکارگیری یادگیری عمیق در حوزه پردازش تصویر و با عنوان شبکه های عصبی کانولوشنی توسط LeCun در سال 1998 انجام شد. در این روش سعی شد که عمل یادگیری و درک تصویر مانند مغز انسان و به صورت چند سطحی انجام شود. شبکه عصبی کانولوشنی معرفی شده در تحقیق نامبرده برای تشخیص اشکال دو بعدی بکار گرفته شده است. الگوریتم معرفی شده در روش های پیشنهادی شبکه مبدل گراف⁵ نام داشته و برای تشخیص کاراکترهای دست نویس به کار رفته است.

1-2- اهمیت و ضرورت پایان نامه

1 Shallow Learning

2 Speech Recognition

3 Computer Vision

4 Natural Language Processing

5 Graph Transformer Network

در طول چند دهه اخیر و همزمان با گسترش و پیشرفت زبان‌شناسی محاسباتی، در بسیاری از کشورهای پیشرفته، تلاش‌های همه‌جانبه و پیگیر در جهت خلاصه‌متون به کمک کامپیوتر انجام گرفته است. تحقیقات بر روی زبان، متن و پردازش بر روی آنها در سالهای اخیر و به خصوص در انجمن پردازش سیگنال گفتار و متن **IEEE** بسیار مورد توجه قرار گرفته است. کاربرد یادگیری عمیق در این حوزه با مدل‌های زبانی شروع شد. هدف مدل‌های زبانی یافتن احتمال وجود یک دنباله از کلمات یا سایر سمبول‌های زبانی مانند حروف، کاراکترها و یا آوای زبانی در متن است. از این رو به کارگیری یادگیری عمیق به عنوان یکی از مهمترین و فعال‌ترین حوزه‌های تحقیقاتی در پردازش زبان طبیعی، می‌تواند در حوزه پردازش متن نیز بسیار موثر واقع شود. بیشتر کارهای انجام شده توسط یادگیری عمیق در حوزه پردازش متن شامل آموزش بردار کلمات توسط مدل‌های زبانی و ترکیب این روش‌ها با سایر ویژگی‌های قابل استخراج برای طبقه‌بندی است.

1-3- اهداف مشخص پایان نامه

در این پژوهش تلاش بر این است با مطالعه روش‌ها و مدل‌های خلاصه‌سازی متن الگوهایی پیشنهاد شود که پس از استفاده از این الگوها علاوه بر افزایش سرعت و دقت، متن خلاصه شده متنی منسجم با معنا و مفهومی هرچه نزدیک‌تر به اصل متن در زبان مبدأ باشد.

1-4- سئوالات تحقیق

سوالاتی که در این تحقیق به آنها پاسخ داده میشود از این قرار است:

- 1) هوش مصنوعی چیست؟
- 2) یادگیری ماشینی^۱ به چه معنایی است؟
- 3) شبکه‌های عصبی مصنوعی چیست؟
- 4) انواع شبکه‌های عصبی مصنوعی کدامند؟
- 5) یادگیری عمیق چیست؟
- 6) الگوریتم‌های یادگیری عمیق کدامند؟
- 7) معماری عمیق در یادگیری کدامند؟
- 8) کاربردهای معماری عمیق چیست؟

1-5- تعریف واژه‌ها و اصطلاحات فنی و تخصصی

1-5-1 یادگیری عمیق

یادگیری عمیق شاخه ای از رشته یادگیری ماشین است که هدف اصلی یادگیری ماشین که همان هوش مصنوعی است را معرفی می کند. یادگیری عمیق تقریباً شامل یادگیری سطوح چند گانه از نمایش و چکیدگی برای درک بهتر از داده هایی مانند: عکس، صدا و متن است. یادگیری عمیق در واقع اشتراک زمینه های تحقیقاتی شبکه های عصبی، هوش مصنوعی، مدل سازی گرافیکی، بهینه سازی، تشخیص الگو و پردازش سیگنال می باشد.

1-5-2 شبکه های عصبی مصنوعی

شبکه عصبی مصنوعی یک سامانه پردازشی داده ها است که از مغز انسان ایده گرفته و پردازش داده ها را به عهده پردازنده های کوچک و بسیار زیادی سپرده که به صورت شبکه ای به هم پیوسته و موازی با یکدیگر رفتار می کنند تا یک مسئله را حل کنند. در این شبکه ها به کمک دانش برنامه نویسی، ساختار داده ای طراحی می شود که می تواند همانند نرون¹ عمل کند. بعد با ایجاد شبکه ای بین این نرونها و اعمال یک الگوریتم آموزشی به آن، شبکه را آموزش می دهند.

یک شبکه عصبی مصنوعی (ANN)² ایده ای است برای پردازش اطلاعات که از سیستم عصبی زیستی الهام گرفته شده و مانند مغز به پردازش اطلاعات می پردازد. عنصر کلیدی این ایده، ساختار جدید سیستم پردازش اطلاعات است. این سیستم از شمار زیادی عناصر پردازشی فوق العاده بهم پیوسته تشکیل شده که برای حل یک مسأله با هم هماهنگ عمل می کند.

1-5-3 شبکه های کانولوشن

شبکه های عصبی کانولوشن تا حد بسیار زیادی شبیه شبکه های عصبی مصنوعی هستند. این نوع شبکه ها متشکل از نرون هایی با وزن ها و بایاس های قابل یادگیری (تنظیم) هستند.

1-5-4 بینایی ماشین

بینایی ماشین شاخه ای از دانش است که سعی دارد از طریق پردازش تصاویر دوبعدی، جهان سه بعدی پیرامون

1 neurons

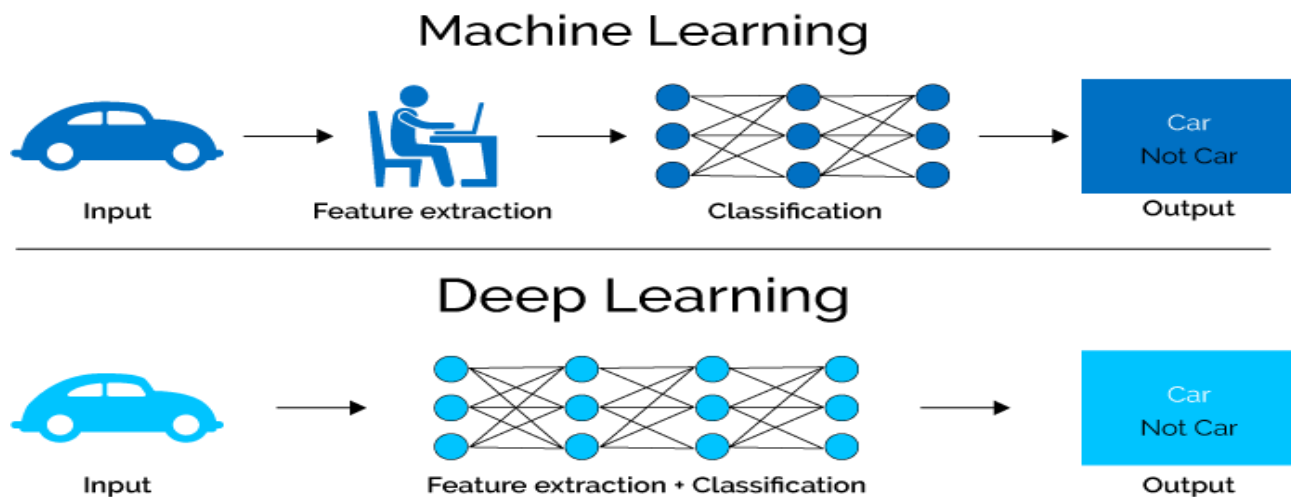
2 Artificial Neural Network

را بازسازی و تفسیر کند. به بیان ساده، بینایی ماشین یعنی اینکه کامپیوترها بتوانند جهان را به کمک دوربین‌ها ببینند، بفهمند و حتی از بینایی انسان پیشی بگیرند.

1-5-5 یادگیری ماشینی¹

یادگیری ماشینی مجموعه‌ای از تکنیک مورد استفاده برای پردازش داده‌های بزرگ با توسعه الگوریتم‌ها و مجموعه‌ای از قوانین برای ارائه دادن نتایج موردنیاز به کاربران محسوب می‌شود. این روش بیشتر برای توسعه ماشین‌های اتوماتیک با اجرا الگوریتم و مجموعه‌ای از قوانین از پیش تعریف شده به کار گرفته می‌شود.

در یادگیری ماشینی داده‌ها مورد استفاده قرار می‌گیرند و مجموعه‌ای از قوانین توسط الگوریتم مورد اشاره اجرا خواهد شد. از این رو تکنیک‌های یادگیری ماشینی به عنوان دستورالعمل‌هایی به شکل خودکار جهت تولید نتایج مطلوب مورد طبقه بندی قرار می‌گیرند. البته این عمل بدون نقش انسان قابل انجام بوده و به شکلی خودکار داده را به الگو تبدیل می‌کند و به شکل اتوماتیک مشکل تولید را شناسایی خواهد کرد.



1-5-6 شبکه‌های عصبی زیستی

شبکه‌های عصبی زیستی مجموعه‌ای بسیار عظیم از پردازشگرهای موازی به نام نورون اند که به صورت هماهنگ برای حل مسئله عمل می‌کنند و توسط سیناپس‌ها (ارتباط‌های الکترومغناطیسی) اطلاعات را منتقل می‌کنند. در این شبکه‌ها اگر یک سلول آسیب ببیند بقیه سلول‌ها می‌توانند نبود آنرا جبران کرده و نیز در بازسازی آن سهیم باشند. این شبکه‌ها قادر به یادگیری اند. مثلاً با اعمال سوزش به سلول‌های عصبی لامسه، سلول‌ها یاد می‌گیرند که به طرف جسم داغ نروند و با این الگوریتم سامانه می‌آموزد که خطای خود را اصلاح کند. یادگیری در این سامانه‌ها به صورت تطبیقی صورت می‌گیرد، یعنی با استفاده از مثال‌ها وزن سیناپس‌ها به گونه‌ای تغییر می‌کند که در صورت دادن ورودی‌های تازه سامانه پاسخ درستی تولید کند.

1-5-7 پردازش زبان های طبیعی¹

پردازش زبان های طبیعی به عنوان زیرمجموعه‌ای از هوش مصنوعی، می‌تواند توصیه‌ها و بیانات را با استفاده از زبانی که شما به طور طبیعی در مکالمات روزمره بکار می‌برید، بفهمد و مورد پردازش قرار دهد. به طور کلی نحوه کار این شاخه از هوش مصنوعی این است که زبانهای طبیعی انسان را تقلید می‌کند. در این میان، پیچیدگی انسان از بعد روانشناسی بر روی ارتباط متعامل تاثیر می‌گذارد. در پردازش زبانهای طبیعی، انسان و کامپیوتر ارتباطی کاملاً نزدیک با یکدیگر دارند. کامپیوتر از لحاظ روانی در مغز انسان جای داده میشود. بدین ترتیب یک سیستم خلاق شکل می‌گیرد که انسان نقش سازمان دهنده اصلی آن را برعهده دارد. اگرچه هنوز موانع روانشناسی و زبانشناسی بسیاری برسر راه سیستم‌های محاوره‌ای وجود دارد. اما چشم اندازهای پیشرفت آنها یقیناً نویدبخش است. در حقیقت، توقعات یکسان از محاوره انسان_ماشینی و محاوره انسان_انسان، معقول نیست.

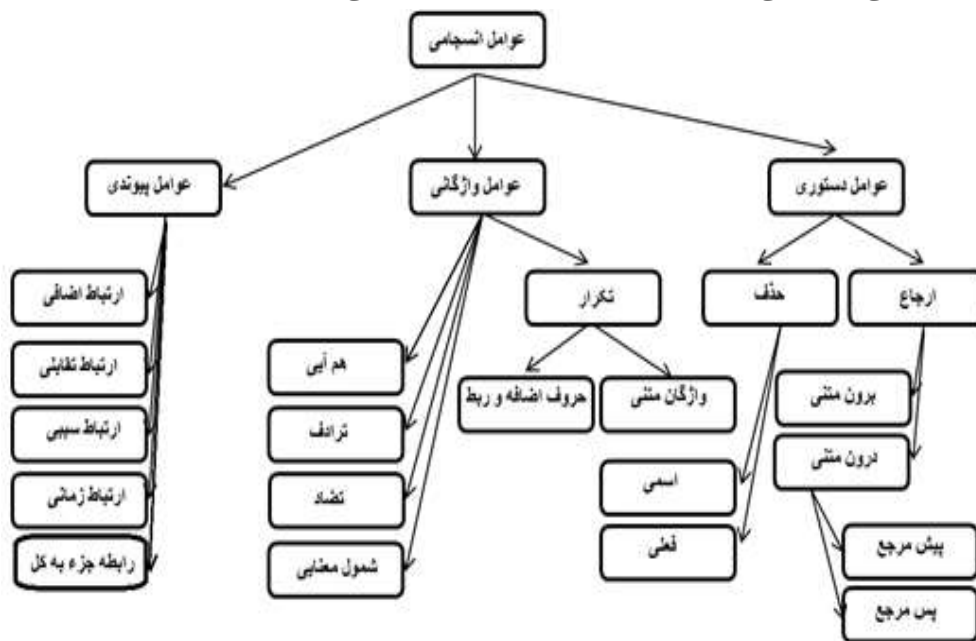
1-5-8 پردازش و انسجام متن

متنهای غیر منسجم اغلب خروجی سیستمهای پردازش متن هستند. میتوان رویکردهای متفاوت ارزیابی انسجام را بر اساس این سیستمها دسته‌بندی کرد. یک متن غیر منسجم میتواند خروجی سیستمهای تولید کننده متن، خلاصه‌سازی، ساده‌سازی، امتیازدهی خودکار مقالات، متن تولید شده توسط یک سیستم پرسش و پاسخ و یا حتی متن تولید شده توسط یک فرد اما با دانش نگارشی پایین باشد. لذا در تمامی سیستمهای پردازشی ذکر شده ارزیابی انسجام و پیوستگی موضوعی متن خروجی، مهمترین بخش میباشد. به عنوان مثال برای یک خواننده متن خلاصه شده، انسجام و پیوستگی آن نیز به همان اندازه جامعیت و کامل بودن مفاهیم اصلی موجود در نوشته اصلی اهمیت دارد. تولید متن منسجم، از همان ابتدای معرفی خلاصه‌سازی ماشینی

توسط Luhn در سال ۱۹۵۸ مورد توجه قرار گرفت [5-1]. اما امروزه با بزرگتر شدن متنها، امکانات جستجوی قویتر، دستیابی و ترکیب خیل عظیم منابع متنی در وب، ایجاد رویکردهای پیشرفته خلاصه‌سازی، تولید متن و سایر عملیات پردازشی، انسجام و پیوستگی متون تولید شده مهمترین دغدغه در این سامانه‌ها می‌باشد. به عنوان مثال متن خلاصه شده توسط یک سیستم خلاصه‌سازی استخراجی از انسجام بسیار کمی برخوردار است. از آنجا که در خلاصه‌های استخراجی تنها معیار ایجاد یک متن فشرده تشخیص و استخراج جملات مهم آن می‌باشد، با حذف برخی از جملات کم اهمیت‌تر انسجام متن خروجی کاهش یافته، به صورتی که خواننده حتما متوجه خلا ایجاد شده بین جملات موجود در خلاصه خواهد شد. بنابراین میتوان اینگونه نتیجه گرفت برخی از جملات حذف شده با وجود اهمیت کم در الگوها استخراجی، ممکن است دارای الگوهای انسجامی بالایی بوده و نباید حذف شوند. در ادامه به معرفی دسته‌بندی رویکردهای متفاوت ارزیابی انسجام از این دیدگاه پرداخته میشود.

1-5-9 پارامترهای مشخص کننده یک متن منسجم

مناسبترین عوامل و ویژگی‌های موجود در یک متن جهت ارزیابی انسجام آن با رویکردهای آماری عوامل واژگانی است. اما معرفی سایر ویژگی‌های معنایی قابل استخراج از متن نیز خالی از لطف نیست. با توجه به اینکه تاکنون تحقیقات زیادی برای تعیین و مشخص کردن میزان انسجام یک متن در زبان‌های مختلف صورت گرفته است، اما هنوز فقط به بخشی از ویژگی‌های ذکر شده پرداخته شده و بخشی از آنان به کار گرفته نشده است.



عوامل انسجامی معرفی شده توسط هالیدی

1- عوامل دستوری

ارجاع: تعبیر و تفسیر عنصری در متن وابسته به تعبیر و تفسیر عنصری دیگر در بخش دیگری از متن باشد. این ارتباط به دو صورت است:

❖ درون متنی: هر دو عنصر در یک متن و در جملات تقریباً نزدیک به هم بوده که می توانند به دو صورت باشند:

✓ پیش مرجع: جمله شامل عنصر مرجع پیش از جمله شامل عنصر مرجوع قرار دارد.

✓ پس مرجع: جمله شامل عنصر مرجع پس از جمله شامل عنصر مرجوع قرار دارد.

❖ برون متنی: دو عنصر در متنهای موازی با هم بوده که در ویژگی بینامتنی مورد توجه قرار گرفته و به انسجام تعدادی از متون با هم مربوط می شود.

حذف: منظور حذف یک یا چند عنصر در جمله در قیاس با عناصر قبل در متن بوده و شامل موارد زیر میباشد:

✓ اسمی

✓ فعلی

2- عوامل واژگانی

تکرار: منظور از این ارتباط تکرار عینی عناصری از متن در جملات متوالی بوده و به دو صورت است:

❖ واژگان متنی: تکرار عینی یک واژه یا عبارت در جملات متوالی و نشانی از ارتباط انسجامی بین این جملات است. این نوع واژگان در صورتی که فاعل یا مفعول باشند میتوانند در جملات بعدی در جایگاه ضمیر قرار گیرند.

❖ حروف اضافه و ربط: هر چند این عناصر از مهمترین عوامل انسجام و پیوستگی درون هر جمله‌ای بوده، اما نمیتوانند دلیلی بر انسجام بین جملات متن باشند. ایست‌واژه‌ها بخشی از این عناصر هستند.

هم‌آیی: نوع دیگری از انسجام واژگانی بوده و به مفهوم با هم آمدن کلماتی در چارچوب موضوع متن است. این واژگان میتوانند به یک حوزه معنایی تعلق داشته و یا مکمل هم باشند. بهترین مثال از این نوع n-gram ها هستند. اگر دو یا چند واژه از این نوع در جملات مجاور با هم قرار داشته باشند دلیلی بر پیوستگی و انسجام آنان خواهند بود.

ترادف: وجود کلمات هم معنی و مترادف در جملات متوالی میتواند به مفهوم ارتباط انسجامی بین آنها باشد.

تضاد: گاهی اوقات تضاد در مفهوم و معنی دو یا چند کلمه یا عبارت در بخشهایی از متن میتواند به مفهوم ارتباط انسجامی بین آن دو بخش شود.

شمول معنایی: این ارتباط به این معنی است که یک مفهوم بتواند چند مفهوم دیگر را شامل شود. مانند رابطه بین مفهوم درخت و واژه های کاج، صنوبر، ...

جزء به کل: این رابطه ارتباط معنی بین دو مفهوم کل و جزء را در بین دو بخش از متن مشخص میکند. مانند بخشی از متن که در مورد دست صحبت کرده و بخشی دیگر در مورد انگشتان دست سخن به میان آورد.

عوامل پیوندی: این عوامل شامل روابط معنایی و منطقی میان جملات سازنده یک متن بوده و شامل اجزای زیر است:

ارتباط اضافی: این رابطه معنایی زمانی برقرار میشود که جمله‌ای در باره محتوای جمله یا جملات قبلی خود موضوع یا مطلبی را اضافه کند. مطالب افزوده شده میتوانند توضیحی، تمثیلی و یا مقایسه‌ای باشند.

ارتباط تقابلی: این رابطه زمانی برقرار شده که محتوی یک جمله خلاف انتظاراتی باشد که جمله یا جملات قبلی در مخاطب بوجود آورده است. از نشانه های متنی این رابطه میتوان به اما، ولی، اگر چه، با وجود این، ... اشاره کرد. به این ارتباط در زبان انگلیسی Shift of idea گفته میشود. مثلا: "علی در مهر و آبان در درس ریاضی بسیار ضعیف عمل کرده است. این روند درس خواندن حتما موجب عدم موفقیت او خواهد شد. اما اخیرا با تلاش بیشتری برای جبران عقب ماندگی خود بر روی این درس کار میکند".

ارتباط سببی: این رابطه نیز هنگامی در یک متن برقرار میشود که یک رویداد در یک جمله به رویدادی در جمله دیگر ارتباط برقرار کند. ارتباط سببی شامل روابطی مانند علت، شرط، نتیجه و هدف میباشد. از نشانه‌های این نوع ارتباط در زبان فارسی میتوان به کلماتی مانند: زیرا، برای اینکه، بنابراین، لذا، به علت، به خاطر، از آنجایی که، چون، در نتیجه، ... اشاره کرد.

ارتباط زمانی: این رابطه زمانی بین دو جمله برقرار میشود که گونه‌ای از توالی زمانی بین رویدادهای دو جمله وجود داشته باشد. جملات موجود در متنهایی مانند زندگینامه، فیلمنامه، داستان و متنهای روایی دارای بیشترین ارتباط زمانی میباشند. ارتباط زمانی جملات میتواند کمک شایانی در بخش جایگذاری جملات استخراجی خلاصه شده در خروجی بکند. [1]

10-5-1 یادگیری عمیق و بکارگیری آن در پردازش متن

تحقیقات بر روی زبان، متن و پردازش بر روی آنها در سال های اخیر و به خصوص در انجمن پردازش سیگنال گفتار و متن IEEE¹ بسیار مورد توجه قرار گرفته است. کاربرد یادگیری عمیق در این حوزه با مدل های زبانی آغاز شد. هدف مدل های زبانی یافتن احتمال وجود یک دنباله از کلمات یا سایر سمبول های زبانی مانند حروف، کاراکترها و یا آواهای زبانی در متن است. از این رو به کارگیری یادگیری عمیق به عنوان یکی از مهم ترین و فعال ترین حوزه های تحقیقاتی در پردازش زبان طبیعی، می تواند در حوزه پردازش متن نیز بسیار موثر واقع شود. البته لازم به ذکر است که بکارگیری یادگیری عمیق در حوزه پردازش متن به اندازه شناسایی گفتار و بینایی ماشین پیشرفت نکرده است. این عدم پیشرفت به دلیل عدم معرفی شواهدی برای اثبات برتری یادگیری عمیق نسبت به روش های سنتی در پردازش زبان طبیعی بوده و اکثر محققین این حوزه هنوز با دیده تردید به آن می نگرند. در حالی که در حوزه های پردازش گفتار و بینایی ماشین، برتری استفاده از آن نسبت به روش های قبل با پیشنهاد روش هایی جدید به اثبات رسیده است. بیشتر کارهای انجام شده توسط یادگیری عمیق در حوزه پردازش متن شامل آموزش بردار کلمات توسط مدل های زبانی و ترکیب این روش ها با سایر ویژگی های قابل استخراج برای طبقه بندی است. [2-3-4]

11-5-1 انواع نگارش

عمده روش های نوشتن به این پنج گونه، دسته بندی می شوند: **نمادین**²، **سیلابی**³، **الفبایی**⁴، و **ویژه**⁵، **ایدئولوگ** (هر نشانه برای یک ایده). ششمین گروه نماد تصویری است که به تنهایی برای نشان دادن زبان، کافی نیست ولی معمولاً هسته دسته نمادین را شکل می دهد.

1 Signal Processing Society's Speech and Language Processing Technical Committee.

2 Logographic

3 syllabic

4 alphabetic

5 featural

6 ideographic