

# SID



ابزارهای  
پژوهش



سرویس ترجمه  
تخصصی



کارگاه های  
آموزشی



بلاگ  
مرکز اطلاعات علمی



سامانه ویراستاری  
STES



فیلم های  
آموزشی

## کارگاه های آموزشی مرکز اطلاعات علمی



آموزش مهارت های کاربردی در تدوین و چاپ مقالات ISI

آموزش مهارت های کاربردی  
در تدوین و چاپ مقالات ISI



روش تحقیق کمی

روش تحقیق کمی



آموزش نرم افزار Word برای پژوهشگران

آموزش نرم افزار Word  
برای پژوهشگران



مرکز توسعه فناوری نیرو (متن)



انجمن کامپیوتر ایران  
Computer Society of Iran

## ارزیابی پارامترهای مؤثر و میزان تاثیر آنها در سیستم های خلاصه سازی با استفاده از شبکه عصبی

لیلا شریف  
گروه علوم کامپیوتر  
دانشگاه شهید بهشتی تهران  
[l\\_sharif@sbu.ac.ir](mailto:l_sharif@sbu.ac.ir)

نصرت سلطانیان زاده  
گروه علوم کامپیوتر  
دانشگاه شهید بهشتی تهران  
[s\\_soltanianzadeh@yahoo.com](mailto:s_soltanianzadeh@yahoo.com)

### چکیده

با توجه به افزایش حجم مستندات متنی، برای پاسخگویی به نیازهای اطلاعاتی کاربران، تکنیک های بازیابی اطلاعات به تنهایی کارا نیستند. مطالعه حجم زیاد متون برای کاربران بسیار سخت و زمانگیر است و در اختیار داشتن خلاصه ای از مطالب مهم برای آنها، می تواند بسیار مفید باشد. خلاصه سازی متون باعث استفاده از منابع و مستندات بیشتر با سرعت بالاتر و در نتیجه به دست آمدن اطلاعات با ارزش تر می شود. لذا امروزه استفاده از سیستم های خلاصه ساز از اهمیت زیادی برخوردار است.

برای خلاصه سازی سه نگرش وجود دارد: نگرش سطحی و نگرش عمیق تر و نگرشی که ترکیبی از این دو است. در نگرش سطحی خلاصه تولید شده شامل جملاتی است که تنها از متون استخراج شده است در حالیکه در نگرش عمیق تر جمله ها برگرفته از متون اولیه هستند. از آنجا که طراحی سیستم هایی که خلاصه ای با کیفیتی مشابه با خلاصه انسانی تولید کنند کار دشواری است محققان از تکنیک های استخراج جمله استفاده می کنند که یک نوع از خلاصه سازی می باشد.

برای انتخاب جمله ها از متن اولیه پارامترهایی برای هر جمله تعریف می شود. در بررسی های به عمل آمده در گذشته اهمیت این پارامترها در استخراج جمله یکسان در نظر گرفته می شده است. حال آنکه ممکن است تاثیر یکی از دیگری در متون مختلف متفاوت باشد. هدف ما در این تحقیق ارائه روشی برای تعیین اهمیت هر یک از این پارامترها در انتخاب جملات می باشد. شبکه عصبی روشی است که می توان برای این امر از آن استفاده کرد. همچنین می توان با استفاده از شبکه عصبی دریافت که جملات انتخابی چه ویژگیهایی باید داشته باشند. با آموزش شبکه عصبی پارامترها و میزان اهمیت آنها در انتخاب جملاتی که باید در خلاصه باشند تعیین می شود.

### کلمات کلیدی

خلاصه سازی متن<sup>۱</sup>، استخراج جمله، شبکه عصبی پیشرو<sup>۲</sup>، الگوریتم پس انتشار خطا<sup>۳</sup>.

### ۱. مقدمه

امروزه با رشد سریع اطلاعات و داده ها، یافتن اطلاعات مناسب و کارا از اهمیت خاصی برخوردار است. هدف از خلاصه سازی متن، فراهم کردن خلاصه ای از محتویات مطابق با اطلاعات مورد نیاز کاربر است. بنابراین کاربران می توانند با سرعت بیشتری به اطلاعات موجود در متن دسترسی پیدا کنند. خلاصه سازی در اصل فرآیند فشرده سازی یک منبع است طوری که نتیجه کار شامل اطلاعات مهم آن منبع باشد. به عبارتی استخراج قسمتهای مهم از یک یا چند متن را خلاصه سازی گویند.

به طور کلی خلاصه سازی می تواند به دو صورت انجام شود:

- خلاصه سازی استخراجی (Extract): جملات مهم به همان صورتی که در متن اصلی بیان شدند شناسایی و لفظ به لفظ در متن خلاصه کپی می شوند.
- خلاصه سازی چکیده (Abstract): جملات متن خلاصه بر گرفته از متن اولیه هستند. به این معنی که جملات به صورت عینی در متن اصلی وجود ندارند.

خلاصه سازی استخراجی ساده ترین روش خلاصه سازی است به طوری که جمله های انتخاب شده از منبع اصلی اطلاعات می باشند. سادگی این روش مسائلی نیز به همراه دارد. یکی از آشکارترین مسائل حفظ حالت اولیه متن است. شکستن متون و نادیده گرفتن اتصال بین جمله

<sup>1</sup> Text summarization

<sup>2</sup> Feed\_forward neural network

<sup>3</sup> Error Backpropagation algorithm

ها می تواند خطرناک باشد. در این روش شانس کمی وجود دارد که اطلاعات مهم از بین نروند [۱].

نوع دیگری از خلاصه سازی استخراج الگوها و تولید چکیده است. در این روش که خلاصه تولید شده چکیده نامیده می شود ممکن است جملات در متن اصلی وجود نداشته باشند. در حقیقت این متن استنتاجی از متن اولیه است. از آنجا که برای ایجاد چکیده باید متن اصلی را فهمید و درک کرد لذا این نوع خلاصه سازی می تواند از تراکم و فشردگی بیشتری نسبت به خلاصه سازی استخراجی برخوردار باشد. برای استخراج الگوها ابتدا متن اصلی مورد تحلیل و بررسی قرار می گیرد و بعد از ارزیابی مفاهیم با هم ادغام می شوند. این اطلاعات با هم ترکیب شده و تشکیل یک خلاصه می دهند.

تاکنون بیشتر تحقیقاتی که در زمینه خلاصه سازی متون به صورت اتوماتیک انجام شده تولید خلاصه هایی از نوع استخراجی بوده است. اما همیشه این دغدغه که چگونه می توان مهمترین و ارزشمندترین جمله ها را در متون شناسایی کرد وجود دارد. نگرش دیگری که وجود دارد یک نگرش ترکیبی است. به این صورت که می توان تکنیک های مختلف خلاصه سازی را مانند شبکه عصبی و پردازش زبان طبیعی<sup>۴</sup> با هم ترکیب کرد.

در تحقیقات انجام گرفته در سالهای گذشته تکنیک هایی برای خلاصه سازی متون معرفی شده اند. از آن جمله Luhn در سال ۱۹۵۸ استخراج جملات بر اساس تعداد تکرار کلمات (frequency) را مطرح کرد [۲]. در این مطالعات جمله ای که کلمات آن بیشتر تکرار شده باشد به عنوان یک جمله خوب و با ارزش برای متن خلاصه در نظر گرفته می شود. موقعیت جمله ها در متن در سال ۱۹۵۸ توسط Baxendale مورد بررسی قرار گرفت [۳]. بسته به محل قرار گرفتن جمله ها در پاراگراف جمله ها به سه دسته تقسیم می شوند: ابتدای پاراگراف، وسط پاراگراف و انتهای پاراگراف. جملاتی که در ابتدا و انتهای پاراگراف قرار دارند از اهمیت بیشتری نسبت به جملات دیگر برخوردار هستند. Edmundson استفاده از کلمات کلیدی را در سال ۱۹۶۹ پیشنهاد داد [۴]. وجود عباراتی از قبیل "به طور خلاصه" و "در نتیجه" و ... در ابتدای جمله به معنی اهمیت آن جمله است. بعدها استفاده از ویژگیهای زبان شناسی نیز به آنها اضافه شد. طول جمله نیز از نگرش هایی است که به آن توجه می شود. مطالعات انجام شده توسط Kupiec در سال ۱۹۹۵ و Moens و Teufel در سال ۱۹۹۷ نشان می دهد که جملاتی با طول خیلی زیاد یا طول خیلی کوتاه شناسی در انتخاب شدن ندارند.

باید توجه داشت که در نظر گرفتن هر یک از این پارامترها می تواند در کیفیت خلاصه ایجاد شده تاثیرگذار باشد. برای مثال انتخاب پارامتری مثل "خوانایی" می تواند باعث کاهش شباهت جملات به هم در متن خلاصه شود. به همین منظور باید پارامترهایی انتخاب و در سیستم

خلاصه ساز لحاظ شوند که خلاصه تولید شده بیشترین شباهت را با یک خلاصه انسانی داشته باشد.

ما یک نگرش یادگیری ماشین را که از شبکه عصبی برای ارزیابی پارامترها استفاده می کند پیشنهاد می کنیم. یک شبکه عصبی روی مجموعه ای از مستندات آموزش می بیند. شبکه آموزش دیده تغییر یافته و برای مستندات دیگر مورد آزمایش قرار می گیرد. به کمک این شبکه می توان پارامترهای مهم برای تولید خلاصه را مشخص کرد. ادامه این مقاله به صورت زیر سازماندهی شده است:

در بخش ۲ تحقیقاتی که در زمینه خلاصه سازی متون مختلف صورت گرفته و تکنیک های استفاده شده در هر یک از آنها را مرور خواهیم کرد. در بخش ۳ مفاهیمی که در خلاصه سازی مطرح هستند تعریف کرده و به کمک این تعاریفات پارامترهای مؤثر را برای استخراج جمله های موجود در متون در بخش ۴ معرفی می کنیم. در بخش ۵ مرور کوتاهی روی شبکه عصبی و سپس محاسبه میزان تاثیر پارامترها در انتخاب جملات متن توسط یکی از روشهای یادگیری ماشین می پردازیم. روش پیشنهادی از تکنیکهای کلاسیک و آماری برای تولید خلاصه مستخرج از متون بهره می برد. در بخش ۶ چگونگی استفاده از ضرایب به دست آمده را در یک سیستم خلاصه ساز بررسی می کنیم. نتایج حاصل را با نتایج به دست آمده از همین سیستم بدون اعمال ضرایب مقایسه کرده و آن را به صورت جدول و نمودارهایی در بخش ۷ ارائه کردیم. در بخش ۸ نیز یک نتیجه گیری کلی از این کار انجام شده است.

## ۲. پیشینه کار

کارهای انجام شده از سال ۱۹۵۰ تا ۱۹۹۰ را می توان بر اساس نوع نگرش به دو دسته طبقه بندی کرد: نگرشهای آماری و نگرشهای هوش مصنوعی. سیستم هایی که از نگرش های آماری استفاده می کنند سعی در پیدا کردن نشانه هایی دارند که می تواند جمله های مهم و با ارزش را نشان دهد. این نشانه ها می تواند تعداد تکرار کلمات، موقعیت جمله، عبارت های نشانه و یا ساختار نحوی یک جمله باشد. بر اساس این نشانه ها سیستم اهمیت هر جمله را با اختصاص دادن یک امتیاز نشان می دهد. جمله هایی با بالاترین امتیاز به عنوان جمله های با ارزش تر در نظر گرفته شده و استخراج می شوند. در مقابل نگرش های مبتنی بر هوش مصنوعی سعی در فهم معانی جملات با استفاده از تکنیک های پردازش زبان دارد. با تکنیک هایی که برای پردازش زبان وجود دارد فهم عمیق یک متن در صورتی امکان پذیر است که ساختار متن ثابت باشد و یا از الگوهای مشخصی باشد که اغلب استفاده می شود.

سال ۱۹۹۳ را می توان شروع دوره جدیدی از تحقیقات در زمینه خلاصه سازی متن دانست. از سال ۱۹۹۳ به بعد تحقیقات حول چهار نگرش آماری، مبتنی بر دانش، مبتنی بر فهم سطحی و نگرش ترکیبی انجام گرفته است. در سال ۱۹۹۴ Salton و همکارانش از یک مدل فضای برداری در ارزیابی اطلاعات برای اندازه گیری میزان شباهت

<sup>4</sup> Natural language process (NLP)

$$w_{i,j} = tf_{i,j} \times isf_i \quad (3)$$

### ۳\_۲ شباهت جملات با عنوان متن: شباهت هر

جمله با عنوان را می توان به صورت زیر تعریف کرد:

$$Sim(s_j, q) = \frac{\sum_{i=1}^t w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,j}^2} \times \sqrt{\sum_{i=1}^t w_{i,q}^2}} \quad (4)$$

به طوری که

$$w_{i,q} = (0.5 + \frac{0.5 \times freq_{i,q}}{\max_l freq_{l,q}}) \times isf_i \quad (5)$$

$freq_{i,q}$  بیانگر تعداد تکرار عنوان متن در جمله  $i$  ام است.

### ۳\_۳ شباهت دو جمله نسبت به یکدیگر:

شباهت دو جمله به هم به صورت زیر تعریف می شود:

$$Sim(s_m, s_n) = \frac{\sum_{i=1}^t w_{i,m} \times w_{i,n}}{\sqrt{\sum_{i=1}^t w_{i,m}^2} \times \sqrt{\sum_{i=1}^t w_{i,n}^2}} \quad (6)$$

### ۴. انتخاب پارامترها

هر سند شامل مجموعه ای از جملات و هر جمله خود در برگیرنده چند پارامتر می شود. هدف روشهایی که برای انتخاب پارامتر استفاده می شود کاهش پارامترهای ورودی با حذف برخی از پارامترهای نامرتب است [۶]. نوع پارامترهای انتخابی نقش مهمی را در تعیین جملاتی که برای خلاصه انتخاب می شوند بازی می کند. از جمله پارامترهایی که در خلاصه سازی مورد استفاده قرار می گیرد عبارت است از: "کلمات با ارزش"، "طول جمله"، "شباهت به عنوان متن"، "مجموع شباهت ها" و... [۱۳]. پارامترهایی که از آنها در این مقاله استفاده شده است "مجموع شباهت ها"، "طول جمله" و "شباهت به عنوان متن" می باشند. بسته به نیاز کاربر و نوع خلاصه مورد نظر این پارامترها می توانند تغییر کنند و یا پارامتر جدیدی به آنها اضافه شود. در این بخش به تعریف پارامترهای استفاده شده می پردازیم:

جمله ها و پیدا کردن جمله های مهم استفاده کردند. در سال بعد مسئله خلاصه سازی به صورت یک مسئله دسته بندی آماری با استفاده از الگوریتم Bayesian توسط سه محقق به نامهای Kupiec ، Chen و Pedersen مطرح شد. در ۱۹۹۵ McKeown و Radev از نتایج سیستم های استخراج اطلاعات برای ساخت خلاصه ای روان استفاده کردند. Lin و Hovy تکنیک آماری موقعیت جمله را در ۱۹۹۷ به کار بردند. Knight و Marcu در سال ۲۰۰۰ از تکنیک های آماری برای فشرده سازی جملات استفاده کردند. در همان سال Berger و Mittal نیز از تکنیک های مشابهی برای خلاصه سازی صفحات وب بهره بردند. کارهایی با نگرش فهم سطحی توسط Morris و Hirst در سال ۱۹۹۱، Benbrahim و Ahmad در سال ۱۹۹۵، Barzilay و Elhadad در سال ۱۹۹۷ و Baldwin و Morton در ۱۹۹۸ انجام گرفت. Svore در سال ۲۰۰۷ استفاده از تکنیک های یادگیری ماشین را برای خلاصه سازی پیشنهاد داد.

### ۳. مفاهیم مطرح در سیستم های خلاصه ساز

به طور معمول روند اطلاعات در اسناد مختلف یک شکل نیست. به این معنی که بعضی از بخشها از برخی دیگر از اهمیت بیشتری برخوردار هستند. چالش اصلی که در مسئله خلاصه سازی با آن مواجه هستیم تشخیص قسمتهای مهم یک متن از قسمتهای کم اهمیت تر آن است. با این که کارهای زیادی در زمینه تولید خلاصه به صورت چکیده انجام شده ولی بیشتر تحقیقات صورت گرفته در خلاصه سازی یک متن به روش استخراجی بوده است. برای استخراج جمله های مهم از متن ابتدا لازم است تعاریف اولیه ای در این زمینه داشته باشیم:

#### ۳\_۱ سیستم وزن دهی جملات: سیستم های

خلاصه ساز متون اولیه را به تعدادی جمله می شکنند. هر کدام از این جمله ها بر اساس تعداد تکرار کلمات وزنی را به خود اختصاص می دهند. برای محاسبه وزن این جملات از سیستم وزن دهی  $isf\_tf$  استفاده می شود:

$$tf_{i,j} = \frac{freq_{i,j}}{\max_l freq_{l,j}} \quad (1)$$

$$isf_i = \log \frac{N}{n_i} \quad (2)$$

به طوریکه  $tf_{i,j}$  بیانگر تعداد تکرار کلمه  $i$  ام در جمله  $j$  ام و  $isf_i$  بیانگر عکس تعداد تکرار جمله از کلمه  $i$  ام می باشد. در رابطه (۲)،  $N$  تعداد کل جملات و  $n_i$  تعداد جملاتی است که در آن کلمه  $i$  ام وجود دارد.

حال وزن هر جمله به صورت زیر محاسبه می شود:

<sup>5</sup> Thematic word

<sup>6</sup> Sentence length

<sup>7</sup> Topic relation(TR)

<sup>8</sup> Sum similarity (SS)

ساختار اصلی شبکه‌های عصبی مصنوعی بر اساس دو جزء اصلی نرون‌ها و ارتباطات وزن‌دار می‌باشد.

در حال حاضر انواع مختلف شبکه‌های عصبی مصنوعی وجود دارند که عبارتند از: شبکه‌های پرسپترون چند لایه، کوهونن، هاپفیلد و ... که این شبکه‌ها با روش‌های مختلفی آموزش می‌بینند مثل روش انتشار خطا. می‌توان شبکه‌های عصبی را بر اساس شیوه پردازش اطلاعات در آنها، به دو گروه شبکه‌های پیشخور و شبکه‌های بازگشتی تقسیم کرد. در این شبکه‌ها یادگیری از طریق نمونه انجام می‌شود. بدین معنی که مجموعه‌ای از ورودی و خروجی‌های درست به شبکه عصبی داده می‌شود و شبکه عصبی با استفاده از آنها، وزن‌های خود را به گونه‌ای تغییر می‌دهد که در صورت دادن ورودی‌های جدید پاسخ‌های درستی را تولید کند. در این مقاله با استفاده از شبکه‌های عصبی پرسپترون چند لایه پیشخور و الگوریتم آموزشی پس انتشار خطا، میزان اهمیت هر پارامتر را در انتخاب جملات برای متن خلاصه پیش بینی می‌شود.

#### ۵\_۱ استفاده از شبکه عصبی در تعیین ضرایب

طبق تحقیقات به عمل آمده در گذشته اهمیت پارامترها می‌تواند یکسان نباشد. به این معنی که بعضی مواقع تاثیر یکی از آنها از دیگری بیشتر یا کمتر است. میزان اهمیت هر پارامتر بر اساس تابعی به نام تابع fitness محاسبه می‌شود. این تابع به صورت زیر تعریف می‌شود:

$$\text{Fitness} = \frac{\alpha \cdot \text{sumsimilarity} + \beta \cdot \text{TR} + \gamma \cdot \text{lenght}}{\alpha + \beta + \gamma} \quad (12)$$

به طوری که

$$\alpha + \beta + \gamma = 1. \quad (13)$$

$\alpha, \beta, \gamma$  ضرایب اهمیت پارامترهای نامبرده می‌باشد. در گذشته این ضرایب با آزمایش و خطا محاسبه می‌شدند. به این صورت که خروجی تابع هر چه به ۱ نزدیکتر باشد مطلوب تر است. می‌دانیم که این عمل کاری سخت و پرهزینه است. به همین منظور در این مقاله یک روشی بهینه برای محاسبه ضرایب ارائه می‌کنیم. برای این کار از تکنیک شبکه عصبی و الگوریتم‌های مرتبط با آن بهره گرفته ایم.

برای محاسبه ضرایب در مرحله اول شبکه آموزش داده می‌شود تا بتواند نوع جمله‌هایی که باید برای متن خلاصه انتخاب شوند را تشخیص دهد و الگوهای ذاتی یک جمله برای انتخاب شدن یا نشدن را یاد بگیرد. از این شبکه آموزش دیده برای خلاصه کردن متون دیگر و انتخاب جملاتی که ارزش بالاتری دارند استفاده می‌شود. در مرحله بعد از روی اتصالات شبکه میزان اهمیت هر پارامتر به دست می‌آید. مراحل کار شبکه پیشنهادی را می‌توان به طور خلاصه در قالب فلوجارت ۱ نمایش داد:

#### ۴\_۱ مجموع شباهت‌ها: برای محاسبه این پارامتر

نیاز به داشتن شباهت بین هر دو جمله است. ماتریس شباهت بین جمله‌ها را می‌توان به صورت زیر ساخت:

$$\begin{bmatrix} sim_{1,1} & sim_{1,2} & \dots & sim_{1,n} \\ sim_{2,1} & sim_{2,2} & \dots & sim_{2,n} \\ \vdots & \vdots & \dots & \vdots \\ sim_{n,1} & sim_{n,2} & \dots & sim_{n,n} \end{bmatrix} \quad (7)$$

برای محاسبه جمع شباهت‌ها مراحل زیر را باید دنبال کرد:

۱. در ابتدا ماتریس شباهت بین هر دو جمله را محاسبه می‌کنیم.

۲. در هر ردیف از ماتریس مقادیر به صورت زیر با هم جمع می‌شوند:

$$\text{Sim\_sen}_i = \sum_{j=1}^n sim\_matrix[i, j] \quad (8)$$

۳. حال مقدار به دست آمده را نرمال می‌کنیم:

$$\text{Sum Similarity} = \frac{sim\_sen_i}{\max sim\_sen_i} \quad (9)$$

#### ۴\_۲ شباهت به عنوان متن: یک خلاصه خوب دارای

جملاتی است که با عنوان متن اصلی شباهت زیادی دارد. با داشتن رابطه (۴) شباهت جمله‌ها با عنوان به صورت زیر قابل محاسبه است:

$$\text{TR} = \frac{\sum_{i=1}^t sim(s_j, q)}{\max sim(s_j, q)} \quad (10)$$

مقدار این پارامتر در بازه [۰، ۱] قرار می‌گیرد. هر چه شباهت جمله به عنوان متن بیشتر باشد مقدار این پارامتر به ۱ و هر چه شباهت کمتر باشد به صفر نزدیک می‌شود.

#### ۴\_۳ طول جمله: طول جمله می‌تواند بر اساس رابطه زیر

بدرست آید:

$$\text{Length} = \frac{len_i}{\max len_i} \quad (11)$$

$len_i$  تعداد کلمات موجود در جمله  $i$ -ام است.

#### ۵. شبکه عصبی

شبکه‌های عصبی مصنوعی در واقع الهام گرفته از ساختار درهم و توده ای مغز انسان است که در آن میلیونها نرون از طریق ارتباطاتی که با یکدیگر دارند به حل مسائل یا ذخیره سازی اطلاعات می‌پردازند.

### مرحله دوم: مقداردهی اولیه پارامترهای شبکه

از جمله پارامترهای شبکه وزنه‌های اتصالی بین نرون‌ها در لایه‌های مختلف است. این وزنه‌ها در ابتدای کار به صورت تصادفی در بازه بین [۰ و ۱] مقداردهی می‌شوند. سه پارامتر "مجموع شباهت‌ها" و "شباهت به عنوان متن" و "طول جمله" نیز به عنوان ورودی به نرون‌های موجود در لایه ورودی شبکه اعمال می‌شود. ما برای آموزش شبکه و تست آن نیاز به مجموعه‌ای از داده‌های معتبر داریم که از مجموعه مستندات ۲۰۰۲ DUC استفاده می‌کنیم.

یکی دیگر از پارامترهای شبکه تعریف مقدار آستانه‌ای به عنوان حداکثر میزان خطای قابل قبول است. ما در این مقاله این حد آستانه را ۰/۰۰۱ در نظر می‌گیریم.

### مرحله سوم: آموزش شبکه

ما در این گام شبکه عصبی پیشنهادی را با پارامترهایی که در بخش پیشین تعریف شده‌اند به عنوان ورودی و تابع فعال‌سازی موجود در هر نرون (تابع sigmoid) از لایه ورودی به لایه خروجی جلو می‌بریم. این وضعیت را وضعیت استراحت شبکه عصبی گویند. چرا که در این مرحله هیچ تغییری روی پارامترهای قابل تغییر شبکه عصبی اعمال نمی‌شود. با این کار خروجی شبکه به ازای ورودی‌های مختلف محاسبه می‌شود.

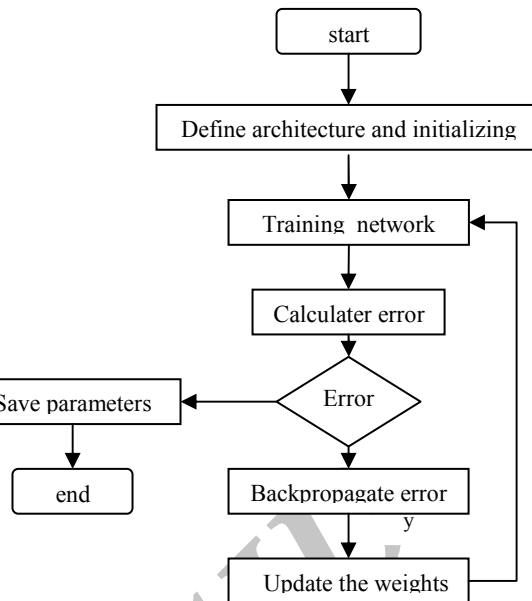
هدف ما از آموزش شبکه عصبی رسیدن به کمترین مقدار تابع خطا است. تابع خطا را می‌توان به صورت زیر تعریف کرد:

$$E = \sum_x E_x = \frac{1}{2} \sum_x (y_x - a_x)^2 \quad (14)$$

به طوری که  $x$  اشاره به  $x$  امین بردار ورودی و  $y_x$  خروجی مطلوب و  $a_x$  خروجی که از شبکه عصبی به دست آمده می‌باشد.

### مرحله چهارم: محاسبه خطا

بعد از به دست آوردن خروجی شبکه برای محاسبه خطا باید آن را با خروجی مطلوب مقایسه کرد. مقصود از خروجی مطلوب وجود یا عدم وجود یک جمله در متن خلاصه است که توسط یک خواننده انسانی تعیین می‌شود. به طوری که این مقدار می‌تواند ۰ یا ۱ باشد. در صورتی که خروجی شبکه همان خروجی مطلوب نباشد باید برای صفر شدن خطا و یا نزدیک شدن این دو خروجی به هم این مقدار را به عقب (به سمت لایه ورودی) برگردانیم و از آن برای تغییر وزنه‌ها استفاده کنیم. اما در صورتی که خطا صفر باشد دیگر نیازی به تغییر وزنه‌ها نیست و میانگین وزنه‌های متصل به هر نرون ورودی که بیانگر یکی از پارامترها است میزان اهمیت آن پارامتر را نشان می‌دهد.

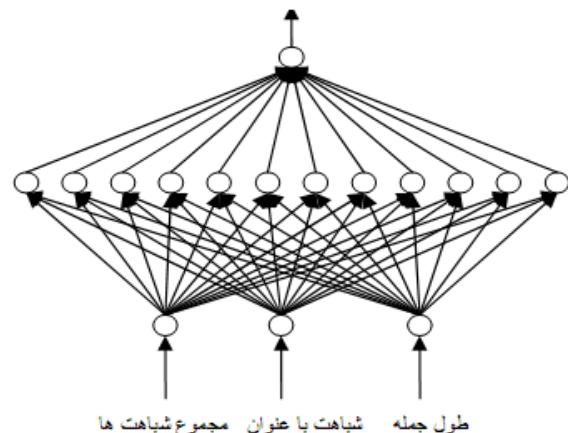


فلوچارت ۱. ساختار شبکه عصبی

### مرحله اول: ساختار شبکه عصبی

ما در این مقاله از یک شبکه عصبی پیشخور سه لایه استفاده می‌کنیم که به عنوان یک تقریب گر برای توابع جهانی نیز به اثبات رسیده است [۷]. شبکه عصبی استفاده شده شامل سه نرون در لایه ورودی و دوازده نرون در لایه میانی و یک نرون در لایه خروجی می‌باشد. این ساختار در شکل ۱ نشان داده شده است. این نرون‌ها در لایه‌های مختلف مرتب شده‌اند به طوری که هر نرون در یک لایه به تمام نرون‌های موجود در لایه بعدی متصل است. آموزش شبکه از نوع آموزش با ناظر است. به این معنی که به ازای ورودی‌های مختلف مقدار عددی مطلوب در دسترس ما قرار دارد. منظور از مقدار مطلوب وجود یا عدم وجود یک جمله در متن خلاصه است که توسط یک خواننده انسانی تعیین شده است.

وجود یا عدم وجود جمله در خلاصه



شکل ۱. ساختار شبکه عصبی

این ضرایب به این معنی است که اهمیت پارامتر "شباهت به عنوان" بیشتر از دو پارامتر دیگر و اهمیت پارامترهای "مجموع شباهت ها" و "طول جمله" با هم برابر می باشد.

#### ۶. استفاده از ضرایب در یک سیستم خلاصه ساز

در این بخش به معرفی یک سیستم خلاصه ساز دو مرحله ای که ضرایب در آن اعمال می شوند می پردازیم. مرحله اول این سیستم در حقیقت یک مرحله پیش پردازش است که روی جملات متن اصلی انجام می شود. بر این اساس، با استفاده از یکسری پارامتر، امتیازی به هر جمله در متن اولیه داده و سپس جملاتی را که امتیازشان از یک حد آستانه<sup>۹</sup> کمتر باشد، حذف می کند. با این کار تعداد جملات متن اصلی کاهش می یابد. در مرحله بعد از الگوریتم جستجوی هارمونی برای تولید خلاصه نهایی استفاده می کند. این سیستم معیارهای دقت و یادآوری را به طور قابل ملاحظه ای افزایش می دهد. ما از این سیستم دو مرحله ای برای ارزیابی ضرایب به دست آمده استفاده می کنیم. در مرحله اول با در نظر گرفتن پارامترهای "مجموع شباهت ها" و "شباهت به عنوان" و "طول جمله" و اعمال ضرایب به دست آمده روی هر کدام طبق رابطه زیر امتیازی به هر جمله در متن می دهیم:

$$\text{score} = (\alpha \times \text{SS}^{10}_{\text{score}} + \beta \times \text{title\_score} + \gamma \times \text{length\_score}) / 3. \quad (18)$$

این مقدار عددی بین ۱۰ و ۱۰۰ است. جملاتی که امتیازشان از حد آستانه (حد آستانه = ۰/۱۵) کمتر است از متن اصلی حذف می شوند. بنابراین با تعداد جملات کمتری روبرو خواهیم بود. سپس در مرحله دوم الگوریتم هارمونی را روی جملات باقی مانده اجرا کرده و متن خلاصه را تولید می کنیم.

#### ۷. ارزیابی

یکی از چالش های اصلی در زمینه خلاصه سازی متون ارزیابی کیفیت خلاصه تولید شده است. معیارهای مختلفی برای تعیین میزان شباهت دو خلاصه وجود دارد. ما در این مقاله از معیارهای دقت و فراخوانی استفاده می کنیم.

$$\text{Precision Rate} = \frac{\text{number of correctly selected sentences}}{\text{total number of selected sentences}} \quad (19)$$

$$\text{Recall Rate} = \frac{\text{number of correctly selected sentences}}{\text{total number of correct sentences}} \quad (20)$$

<sup>9</sup> threshold  
<sup>10</sup> Sum similarity

#### مرحله پنجم: تغییر دادن وزنه های بین هر نرون

زمانی که خطای حاصل از خروجی شبکه و خروجی مطلوب صفر نباشد باید مقدار پارامترهای قابل تغییر شبکه عصبی (وزن های اتصالاتی بین نرون ها در لایه های مختلف) به کمک الگوریتم یادگیری پس انتشار خطا برای کاهش مقدار تابع خطا تغییر کند. این حالت را وضعیت آموزش شبکه عصبی می نامند. با استفاده از الگوریتم نام برده شده پارامترهای متغیر شبکه عصبی که همان وزن های بین نرون ها هستند به صورت زیر تغییر می یابند:

$$w(t+1) = w(t) + \Delta w(t) \quad (15)$$

به طوری که منظور از  $t$  مجموعه وزنه های جاری شبکه می باشد و

$$\Delta w = -\eta \frac{\partial E}{\partial w} \quad (16)$$

$\eta$  نشان دهنده پارامتر نرخ یادگیری است که می تواند به صورت زیر محاسبه شود:

$$\eta = \frac{1.5}{\sqrt{\sum p_i^2}} \quad (17)$$

در این رابطه  $p_i$  بیانگر تعداد کل نمونه هایی که متعلق به دسته  $i$  هستند می باشد.

برای تغییر وزن های لایه خروجی لازم است مشتق تابع خطا را نسبت به وزنه های لایه خروجی به دست آوریم. ما از تابع sigmoid به دلیل محاسبه ساده مشتق آن به عنوان تابع فعال سازی استفاده می کنیم. برای به روز کردن وزن های لایه میانی نیز همانند بالا عمل می کنیم با این تفاوت که مشتق تابع خطا را نسبت به وزنه های لایه میانی محاسبه می کنیم.

بعد از اینکه آموزش شبکه عصبی به ازای تمام نمونه های آموزشی کامل شد می توان از آن برای انتخاب جملات خلاصه از مستندات دیگر بهره گرفت. ما از این شبکه آموزش دیده برای ارزیابی اهمیت هر یک از پارامترهای مطرح شده در قسمتهای پیشین استفاده می کنیم. ضرایب به دست آمده از این شبکه عصبی به ازای پارامترهای "مجموع شباهت ها" و "شباهت به عنوان متن" و "طول جمله" در جدول ۱ نمایش داده شده است:

طول جمله	شباهت به عنوان متن	مجموع شباهت ها	ضریب اهمیت
۰/۲	۰/۶	۰/۲	ضریب اهمیت

جدول ۱. ضرایب به دست آمده از شبکه عصبی

نمودارهایی نمایش داده شده است. این نتایج حاکی از این است که روش پیشنهادی می تواند باعث افزایش کارایی سیستم های خلاصه سازی گردد.

در روابط فوق دقت بیانگر درصد درستی جملات انتخاب شده نسبت به کل جملات انتخاب شده و فراخوانی درصد درستی جملات انتخاب شده نسبت به خلاصه مطلوب (خلاصه انسانی) نشان می دهد. در این مقاله از مجموعه مستندات ۲۰۰۲ DUC به عنوان داده های ورودی برای شبکه عصبی استفاده شده است. برای هر مجموعه از این مستندات خلاصه های انسانی با طول ۲۰۰ و ۴۰۰ کلمه وجود دارد. خلاصه های ایجاد شده با روش پیشنهادی با خلاصه های موجود در این مجموعه مقایسه می شوند. نتایج این مقایسه به صورت جدول ۲ و

جدول ۲. نتایج مقایسه روش پیشنهادی نسبت به الگوریتم هارمونی

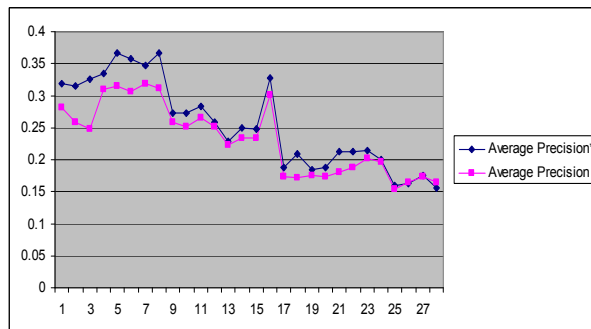
Set No.	Model	Size	نتایج سیستم دو مرحله ای با اعمال ضرایب				نتایج سیستم دو مرحله ای بدون اعمال ضرایب			
			Average Precision*	Average Recall*	Max Precision*	Max Recall*	Average Precision	Average Recall	Max Precision	Max Recall
D061j	B	200	0.31797	0.29813	0.72566	0.64566	0.28147	0.26443	0.63247	0.5873
		400	0.31585	0.29623	0.61111	0.60629	0.259	0.24194	0.56302	0.53174
	I	200	0.32544	0.30565	0.54166	0.51181	0.24842	0.23268	0.53333	0.50793
		400	0.33534	0.32124	0.61764	0.66141	0.30964	0.29527	0.57079	0.53305
D070f	G	200	0.36672	0.29082	0.62992	0.55555	0.31493	0.24316	0.55084	0.45138
		400	0.35774	0.28451	0.64655	0.52083	0.30607	0.23905	0.6129	0.52077
	J	200	0.34725	0.27556	0.63709	0.54861	0.31815	0.24663	0.63157	0.5
		400	0.36572	0.28819	0.75862	0.61111	0.3119	0.24221	0.65	0.54166
D063j	C	200	0.27175	0.24879	0.6055	0.55462	0.2583	0.23453	0.56481	0.5126
		400	0.27291	0.25171	0.5826	0.56302	0.25109	0.22447	0.57522	0.54621
	E	200	0.28305	0.2605	0.5619	0.521	0.26617	0.24357	0.55555	0.5042
		400	0.25887	0.23847	0.57333	0.53305	0.25102	0.22177	0.54285	0.47899
D067f	A	200	0.22799	0.21398	0.72222	0.68421	0.22226	0.21238	0.60416	0.50877
		400	0.2504	0.2379	0.61386	0.5614	0.23403	0.22049	0.59813	0.54385
	I	200	0.24756	0.23679	0.64601	0.64601	0.23415	0.22526	0.58407	0.58407
		400	0.32722	0.30086	0.60189	0.55411	0.30105	0.27718	0.58986	0.54978
D062j	A	200	0.18848	0.19743	0.61224	0.61855	0.17282	0.18494	0.51648	0.48453
		400	0.20954	0.22071	0.57843	0.60824	0.17247	0.18447	0.5567	0.5567
	G	200	0.18431	0.19384	0.5673	0.60824	0.17515	0.18463	0.51647	0.50607
		400	0.1884	0.19659	0.50287	0.50732	0.17288	0.19757	0.50526	0.50607
D113h	A	200	0.21298	0.2049	0.58947	0.53333	0.18012	0.17532	0.56989	0.50476
		400	0.21245	0.20995	0.58333	0.6	0.18782	0.18311	0.56122	0.5238
	I	200	0.21347	0.20981	0.59405	0.57142	0.20196	0.1961	0.56521	0.49523
		400	0.19949	0.1961	0.60396	0.58095	0.19581	0.19264	0.55769	0.55238
D066j	C	200	0.15844	0.15519	0.54125	0.47663	0.1543	0.15024	0.53409	0.43925
		400	0.16271	0.16001	0.51428	0.50467	0.1645	0.15902	0.48543	0.46728
	I	200	0.17493	0.17048	0.49056	0.48598	0.17274	0.16638	0.48113	0.47663
		400	0.15649	0.1525	0.5102	0.46728	0.16394	0.15519	0.50515	0.45794



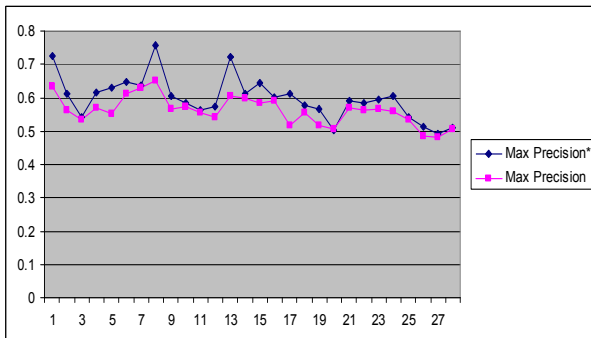
چکیده استفاده می شود. خلاصه سازی برای تولید چکیده کار سختی است اما در مقابل خلاصه سازی استخراجی به راحتی می تواند نتایج رضایتبخش تری در مقیاس های بزرگتر ایجاد کند. این مقاله اشاره به خلاصه سازی استخراجی دارد. انتخاب پارامترها همانند انتخاب جملات برای خلاصه توسط یک خواننده انسانی نقش مهمی را در کارایی شبکه عصبی دارد. شبکه عصبی بر اساس شیوه یک خواننده انسانی و اینکه چه جمله ای را در متون ارزشمند می پندارد آموزش داده می شود. در حقیقت این کار از مزایای نگرش شبکه عصبی است. چرا که هر خواننده می تواند شبکه عصبی را به شیوه خود آموزش دهد. به اضافه اینکه بنابر نیازمندی های مختلف می توان به راحتی پارامترهای انتخابی را تغییر داد.

### منابع

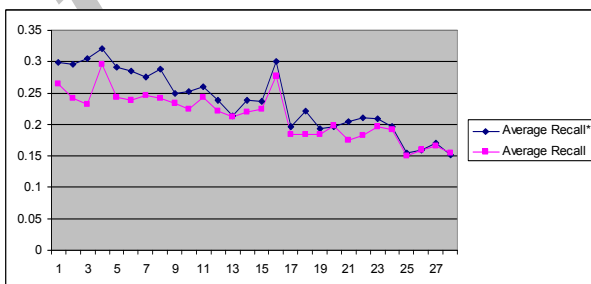
- [1] Almira Mae Diola, Joan Tiffany T. Ong Lopez, Phoebus Ferdiel Torralba, Sherwin So, Allan Borra, "Automatic Text Summarization", College of Computer Studies, De La Salle University.
- [2] Luhn, H. P. (1958). The automatic creation of literature abstracts. IBM Journal of Research Development.
- [3] Baxendale, P. (1958). Machine-made index for technical literature - an experiment. IBM Journal of Research Development.
- [4] Edmundson, H. P. (1969). New methods in automatic extracting. Journal of the ACM.
- [5] J. Kupiec, J. Pederson and F. Chen, "A Trainable Document Summarizer", Proceedings of the 18<sup>th</sup> Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, Washington 1995.
- [6] Forman, G., An Experimental Study of Feature Selection Metrics for Text Categorization. Journal of Machine Learning Research, 2003.
- [7] M.R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems", Journal of Research of the National Bureau of Standards, 1952.
- [8] Svore, K., Vanderwende, L., and Burges, C. (2007). Enhancing single-document summarization by combining RankNet and third-party sources. In Proceedings of the EMNLP-CoNLL.
- [9] Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N., and Hullender, G. (2005). Learning to rank using gradient descent. In ICML '05: Proceedings of the 22nd international conference on Machine learning.
- [10] Radev, D. R., Hovy, E., and McKeown, K. (2002). Introduction to the special issue on summarization Computational Linguistics.



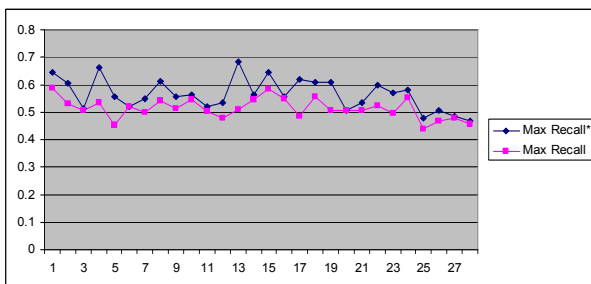
نمودار ۱. مقایسه average precision با اعمال و عدم اعمال ضرایب



نمودار ۲. مقایسه max precision با اعمال و عدم اعمال ضرایب اهمیت



نمودار ۳. مقایسه average recall با اعمال و عدم اعمال ضرایب اهمیت



نمودار ۴. مقایسه max recall با اعمال و عدم اعمال ضرایب اهمیت

### ۸. نتیجه

با توجه به افزایش حجم مستندات، توسعه سیستم های خلاصه ساز که از دقت و کارایی بالایی برخوردار باشند امری لازم است. هر چند از دهه ۵۰ تحقیقات در این زمینه انجام می شود اما هنوز در آن جای کار وجود دارد. بسته به نوع کاربرد از خلاصه های استخراجی و

[11] T. M. Mitchell. Machine Learning. McGRAW-HILL,1995.An Introduction to Back-Propagation Neural Networks, Pete McCollum.

[12] Haykin S. Neural networks: a comprehensive foundation. 2nd ed. New York: Prentice-Hall; 1999.

[13] Joel Larocca Neto, Alex A. Freitas, Celso A. A. Kaestner, Automatic Text Summarization using a Machine Learning Approach Pontifical Catholic University of Parana (PUCPR) Rua Imaculada Conceicao, 2000.

Archive of SID

# SID



ابزارهای  
پژوهش



سرویس ترجمه  
تخصصی



کارگاه های  
آموزشی



بلاگ  
مرکز اطلاعات علمی



سامانه ویراستاری  
STES



فیلم های  
آموزشی

## کارگاه های آموزشی مرکز اطلاعات علمی



تازه های آموزش  
آموزش مهارت های کاربردی در تدوین و چاپ مقالات ISI

آموزش مهارت های کاربردی  
در تدوین و چاپ مقالات ISI



تازه های آموزش  
روش تحقیق کمی

روش تحقیق کمی



تازه های آموزش  
آموزش نرم افزار Word برای پژوهشگران

آموزش نرم افزار Word  
برای پژوهشگران