Arcl

JCWR2020

کنفرانس بین المللی وب پژوهی
ایران، تهران، ۲۲ خرداد ۱۳۹۹

Multi-document Online News Summarization by Learnable Submodular Functions

Alireza Ghadimi [1]*, Hamid Beigy [2]

[1] Research Institute for Information and Communication Technology, Academic Center for Education, Culture and Research (ACECR), Tehran, Iran
ghadimi@ictrc.ac.ir

[2] Computer Engineering Faculty, Sharif University of Technology, Tehran, Iran
beigy@sharif.edu

Abstract

With the increasing rate of online information generation and the growth of Internet users, the need for automatic text summarization has become more urgent. In this paper, a multi-document summarization method based on learnable submodular functions is provided. Using submodularity, it is possible to guarantee the quality of solution. Input text documents are modeled using weighted graphs, where nodes and edges represent sentences and similarities between them, respectively. Using this graph, the features that indicate the significance and impact of each sentence are extracted. Significance features consider the value of any sentence, independent of other ones. Modular functions are used to model these features. On the other hand, impact features consider the relationships between sentences. For modeling these features, submodular functions are used. These features are the building blocks of the target function which is equivalent to an ordinary neural network. Therefore, using a training set, network learns how to summarize text. After learning phase, the obtained function is used to summarize input texts. The summarizer has been examined using Pasokh and DUC 2004 datasets, and its results are presented.

Keywords: Multi-Document Summarization, Submodularity, Neural Network , Weighted Graph.

www.SID.ir

Arcl

ICWR2020
کنفرانس بین المللی وب پژوهی
ایران، تهران، ۲۲ خرداد ۱۳۹۹

خلاصه‌سازی چندسندی اخبار برخط مبتنی بر توابع زیرپیمانه با قابلیت یادگیری

علیرضا قدیمی[۱]*، حمید بیگی[۲]

[۱] پژوهشکده فناوری اطلاعات و ارتباطات جهاد دانشگاهی، تهران، ایران
ghadimi@ictrc.ac.ir

[۲] دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف، تهران، ایران
beigy@sharif.edu

چکیده

با افزایش انتشار برخط اطلاعات و گسترش کاربران اینترنت، نیاز به خلاصه‌سازی خودکار متن ضرورت بیشتری یافته است. در این مقاله، یک روش خلاصه‌سازی چند سندی مبتنی بر توابع زیرپیمانه قابل یادگیری ارایه می‌شود. با استفاده از زیرپیمانگی، امکان تضمین کیفیت راه حل وجود دارد. در روش ارایه شده، سندهای متنی ورودی با استفاده از گراف‌های وزن‌دار مدل‌سازی می‌شوند که راس‌ها، بازنمایی‌کننده‌ی جملات و یال‌ها نشان‌دهنده‌ی شباهت بین جملات است. با استفاده از این گراف، ویژگی‌هایی که نشان‌دهنده‌ی اهمیت و تاثیر هر جمله است، استخراج می‌شوند. ویژگی‌های اهمیت، ارزش هر جمله را مستقل از جملات دیگر در نظر می‌گیرند. برای مدل‌سازی این ویژگی‌ها، از توابع پیمانه‌ای استفاده می‌شود. ویژگی‌های تاثیر، ارزش هر جمله را با توجه به ارتباط آن با جملات دیگر در نظر می‌گیرند. برای مدل‌سازی این ویژگی‌ها، از توابع زیرپیمانه‌ای استفاده می‌گردد. این ویژگی‌ها اجزای سازنده‌ی تابع هدف را تشکیل می‌دهند که معادل با یک شبکه‌ی عصبی متعارف است. از همین رو، از یک مجموعه‌ی آموزشی جهت آموزش شبکه استفاده می‌شود. پس از اجرای فرآیند یادگیری، از این تابع به منظور خلاصه‌سازی استخراجی چندسندی استفاده می‌شود. این خلاصه‌ساز با استفاده از پیکره‌ی پاسخ و DUC 2004 آزمایش شده، و نتایج حاصل از آن ارایه گردیده است.

کلمات کلیدی

خلاصه‌سازی چندسندی - زیرپیمانگی - شبکه عصبی - گراف وزن‌دار

۱- مقدمه

با افزایش انتشار برخط اطلاعات و گسترش کاربران اینترنت، نیاز به خلاصه‌سازی متن بیش از پیش مورد توجه قرار گرفته است. رشد پرشتاب فناوری‌های اطلاعاتی و ارتباطاتی سبب تولید حجم گسترده‌ای از اسناد الکترونیکی شده است که افزونگی دارند و کاربران را برای یافتن اطلاعات مرتبط در راستای نیازهای خود، با چالش روبه‌رو می‌سازند. از این رو ممکن است کاربر از بخش‌های مهم و اصلی اسناد متنی غافل بماند. این چالش، ضرورت نیاز به یک خلاصه‌ساز متن را آشکار می‌سازد. اهمیت خلاصه‌سازی خودکار متن مواردی مانند کاهش زمان مطالعه، افزایش کارآمدی نمایه‌گذاری و سوگیری کمتر نسبت به خلاصه‌سازی توسط انسان را در بر می‌گیرد [1].
خلاصه‌سازی کاربردهای متنوعی دارد. یکی از اساسی‌ترین این کاربردها، خلاصه‌سازی اخبار مبتنی بر وب است [1]. منابع مختلف خبری، اخبار خود در مورد یک رویداد را در پایگاه‌های وبی خود ارایه می‌نمایند و کاربران اینترنت، با مراجعه به این پایگاه‌های وبی، از این اخبار مطلع می‌شوند. با خلاصه‌سازی اخبار منتشر شده حول یک رویداد مشخص، دسترسی کاربران به اخبار تسهیل می‌گردد. این مساله را می‌توان در قالب مساله خلاصه‌سازی چندسندی در نظر گرفت. در این مقاله، با استفاده از راه‌حلی مبتنی بر توابع زیرپیمانه‌ای و یادگیری ژرف به حل این مساله پرداخته می‌شود.
در این روش، سندهای اصلی با استفاده از گراف‌های وزن‌دار مدل‌سازی می‌شوند که راس‌ها، بازنمایی‌کننده‌ی جملات و یال‌ها نشان‌دهنده‌ی شباهت بین جملات است. با استفاده از این گراف، ویژگی‌هایی که نشان‌دهنده‌ی اهمیت و تاثیر هر جمله است، استخراج می‌شوند. منظور از ویژگی‌های اهمیت، ویژگی‌هایی است که ارزش هر جمله را مستقل از جملات دیگر در نظر می‌گیرند. همچنین، ویژگی‌های تاثیر، ارزش هر جمله را با توجه به ارتباط آن با جملات دیگر مورد توجه قرار می‌دهند. این ویژگی‌ها اجزای سازنده‌ی یک

www.SID.ir

تابع زیرپیمانه‌ای هستند که در این مقاله ارائه می‌شود. این تابع معادل یک شبکه‌ی ژرف روبه‌جلو با وزن‌های نامنفی است. از همین رو، از یک مجموعه‌ی آموزشی جهت آموزش شبکه‌ی ژرف استفاده می‌شود. پس از اجرای فرآیند یادگیری، از این تابع به منظور خلاصه‌سازی استخراجی چندسندی استفاده می‌شود. این خلاصه‌ساز با استفاده از پیکره‌ی پاسخ برای زبان فارسی و پیکره‌ی DUC 2004 برای زبان انگلیسی آزمایش شده و نتایج حاصل از آن ارائه گردیده است.

ساختار کلی این مقاله به این صورت است. در بخش دوم، مفاهیم مرتبط با موضوع مقاله مورد بررسی قرار می‌گیرد. در بخش سوم، پژوهش‌های مرتبط ارائه می‌گردد. فصل چهارم، روش ارایه شده تشریح می‌شود و در بخش پنجم، نتایج به دست آمده از آزمایش‌ها ارائه می‌گردد.

۲- مفاهیم مرتبط

برای مجموعه‌ی مرجع V، تابع مجموعه‌ای $f: 2^V \to \mathbb{R}$ را در نظر بگیرید که هر زیرمجموعه‌ی $A \subseteq V$ را به اعداد حقیقی نگاشت می‌دهد. تابع f را نرمال می‌گویند، اگر $f(\phi) = 0$. این تابع را یکنوای غیرنزولی می‌گویند، اگر به ازای هر $A \subseteq B \subseteq V$، رابطه‌ی (۱) برقرار باشد:

$$f(A) \leq f(B) \quad (۱)$$

اگر تابع f نرمال باشد، آنگاه در صورتی پیمانه‌ای است که به ازای هر $A \subseteq V$، رابطه (۲) برقرار باشد [2].

$$f(A) = \sum_{a \in A} f(a) \quad (۲)$$

تابع f را زیرپیمانه‌ای می‌گویند، اگر به ازای هر $A \subseteq B \subseteq V$، رابطه‌ی (۳) برقرار باشد [3]:

$$f(A \cup B) + f(A \cap B) \leq f(A) + f(B) \quad (۳)$$

تعریف هم‌ارز دیگری برای زیرپیمانگی ارائه شده است. در این تعریف، تابع f را زیرپیمانه می‌گویند، اگر برای هر $A \subseteq B \subseteq V$ و $a \in V \backslash B$، رابطه‌ی (۴) برقرار باشد [4]:

$$f(B \cup \{a\}) - f(B) \leq f(A \cup \{a\}) - f(A) \quad (۴)$$

بر اساس رابطه‌ی (۴)، مقدار افزایشی a با رشد زمینه‌ی مورد بررسی از A به B، کاهش می‌یابد. توابع زیرپیمانه‌ای، خصوصیت‌های مشترک بسیاری با توابع مقعر و محدب دارند که از آن جمله می‌توان به گستردگی کاربرد، عمومیت و بسته بودن آنها تحت تعدادی از عملیات ترکیبی متداول مانند جمع، ترکیب توابع و برخی همگشت‌ها اشاره نمود. همان‌گونه که تحدب/تقعر سبب سادگی بهینه‌سازی توابع پیوسته می‌شود، زیرپیمانگی نیز نقشی اساسی را در بهینه‌سازی ترکیبیاتی ایفا می‌نماید [5].

در صورتی‌که تابع هدف یک مساله‌ی بیشنیه‌سازی ترکیبیاتی زیرپیمانه‌ای باشد، با وجود این‌که بیشینه‌سازی توابع زیرپیمانه‌ای یک مساله‌ی ان‌پی-کامل است، می‌توان پاسخی نزدیک به راه‌حل بهینه را در زمان چندجمله‌ای برای آن یافت. یک نتیجه‌ی معروف در این زمینه این است که به کمک یک الگوریتم حریصانه، امکان بیشینه‌سازی یک تابع زیرپیمانه‌ای یکنوا تحت محدودیت اندازه با تقریب مناسبی نسبت به پاسخ بهینه وجود دارد [4].

قضیه ۱، تضمین موجود در مورد ضریب تقریب این الگوریتم حریصانه مطرح می‌نمایند.

قضیه ۱: اگر تابع زیرپیمانه‌ای f، تابعی یکنوای غیر نزولی و نرمال باشد، آنگاه مجموعه‌ی بدست آمده توسط الگوریتم حریصانه، $\hat{A}$، برای بیشینه‌سازی این تابع با محدودیت اندازه‌ی K در بدترین حالت دارای ضریب تقریب برابر با $1 - \frac{1}{e}$ نسبت به مجموعه‌ی بهینه است:

اثبات این قضیه در [6] آمده است.

بر اساس نتایج بدست آمده از قضیه ۱ مشخص می‌شود که با استفاده از یک الگوریتم حریصانه‌ی ساده، امکان ارائه‌ی راه‌حل‌هایی با تقریبی مناسب نسبت به راه‌حل‌های بهینه وجود دارد [7].

بنابراین، الگوریتم حریصانه، یک تقریب ضریب ثابت را در مساله‌ی بیشینه‌سازی زیرپیمانه‌ای با محدودیت اندازه ارائه می‌دهد. نتیجه‌ی بدست آمده از قضیه ۱ دارای اهمیت ویژه‌ای است، زیرا نشان می‌دهد کیفیت راه‌حل به اندازه‌ی مساله وابسته نیست و به همین خاطر حتی برای مسائل بسیار بزرگ هم قابل استفاده است. این موضوع به معنی مقیاس‌پذیری راه‌حل‌های مبتنی بر توابع زیرپیمانه‌ای است. علاوه بر این، یادآوری این نکته لازم است که کران بدست آمده، کران بدترین حالت است و در بسیاری از موارد، کیفیت راه‌حل بدست آمده بسیار بهتر از این کران می‌باشد [5].

در ادامه، قضایایی مطرح شده است که در روش پیشنهادی مورد استفاده قرار گرفته‌اند.

قضیه ۲: اگر $f: 2^V \to \mathbb{R}$ یک تابع زیرپیمانه‌ای غیرنزولی و $g: \mathbb{R} \to \mathbb{R}$ تابعی مقعر و غیرنزولی باشد، ترکیب آنها، $h = gof$ تابعی زیرپیمانه‌ای و غیرنزولی است.

اثبات این قضیه در [5] آمده است.

قضیه ۳: اگر $f: 2^V \to \mathbb{R}_+$، تابعی پیمانه‌ای و نامنفی و $g: \mathbb{R} \to \mathbb{R}_+$ تابعی مقعر، یکنوای غیرنزولی، نرمال و نامنفی باشند، تابع $h = gof$ تابعی زیرپیمانه‌ای است.

اثبات این قضیه در [8] آمده است.

قضیه ٤: خانواده‌ی توابع زیرپیمانه تحت ترکیب خطی نامنفی توابع زیرپیمانه بسته است.

اثبات این قضیه در [2] آمده است.

۳- پژوهش‌های مرتبط

در این بخش، به مرور اجمالی پژوهش‌های مرتبط با خلاصه‌سازی خودکار متن پرداخته می‌شود. کوپیک و همکاران، روشی را ارایه دادند که بر اساس دسته‌بندی بیزین ساده تصمیم می‌گیرد که هر کدام از جملات موجود

www.SID.ir

در سند متنی اصلی، ارزش استخراج برای قرارگیری در خلاصه را دارد یا خیر [9]. اون و همکاران، از دسته‌بند بیز ساده به منظور خلاصه‌سازی استفاده نمودند [10]. آنها در کار خود، از مجموعه‌ی غنی‌تری از ویژگی‌ها بهره بردند و سامانه‌ای را به نام دیم‌سام توصیف کردند که از ویژگی‌هایی مانند فراوانی واژه‌ها و معکوس فراوانی سندها استفاده می‌نمود. کائو و همکارانش، روشی برای رتبه‌بندی بر اساس شبکه‌های عصبی بازگشتی ارائه نموده‌اند که برای خلاصه‌سازی چندسندی قابل استفاده است [11]. در روش ارائه شده، رتبه‌بندی جملات به صورت یک فرآیند رگرسیون سلسله‌مراتبی مدل‌سازی شده است که به صورت همزمان، اهمیت یک جمله و اجزای آن را در یک درخت تجزیه مشخص می‌نماید. در این پژوهش، ابتدا درخت تجزیه‌ی یک جمله استخراج شده و به عنوان ورودی به شبکه عصبی داده می‌شود. شبکه عصبی بازگشتی، اهمیت تمامی گره‌های غیرپایانه‌ی آن جمله را اندازه‌گیری می‌نماید و بر اساس آن، ارزشی را برای کل جمله محاسبه می‌کند. در این پژوهش از ویژگی‌های دست‌ساز مانند فراوانی واژه به عنوان ورودی استفاده می‌شود. ژانگ و همکاران، برای محاسبه‌ی شباهت بین جملات، تشابه واژه‌های موجود در آنها را مورد بررسی قرار دادند و سپس با استفاده از روش خوشه‌بندی کی-مینزس جملات را خوشه‌بندی نمودند [12]. ژولیا و همکارانش، روشی شامل سه گام را برای خلاصه‌سازی مبتنی بر خوشه‌بندی ارائه نمودند. این گام‌ها عبارت هستند از: انتخاب واژه‌ها، وزن‌دهی به واژه‌ها و انتخاب جمله‌ها. در گام نخست، یکی از سه مدل بازنمایی متن، یعنی مدل کیسه‌ی واژه‌ها ، مدل n-گرام و یا مدل دنباله‌ی تکرارشونده‌ی بیشینه برای استخراج واژه‌ها استفاده می‌شود. در گام دوم، واژه‌ها با استفاده از یکی از شاخص‌های فراوانی واژه، معکوس فراوانی سند و یا فراوانی واژه-معکوس فراوانی سند، وزن‌دهی می‌شوند. در گام سوم، با استفاده از الگوریتم بیشینه‌سازی امید ، جملات خوشه‌بندی می‌شوند و یک جمله از هر کدام از خوشه‌ها که بازنمایی‌کننده‌ی هر خوشه است، برای قرارگیری در خلاصه انتخاب می‌گردد . در این روش‌ها تلاش می‌شود تا ابتدا موضوعات مهم سند(ها) استخراج شده و سپس، جملاتی که حاوی این موضوعات هستند برای تشکیل خلاصه انتخاب گردند [13]. در پژوهش دیگری که توسط یوسفی-آذر و همکارانش ارائه گردید، تلاش شده است تا ویژگی‌های سطح بالا از روی ویژگی‌های ورودی یادگیری شوند [14]. در این پژوهش، روشی برای خلاصه‌سازی استخراجی تک‌سندی مبتنی بر پرس و جو با استفاده از خودرمزگذار ژرف ارائه شده است. در این روش، با استفاده از یک خودرمزگذار ژرف، یک نمایش فشرده‌ی سطح بالا از هر جمله به دست می‌آید. سپس برای رتبه‌بندی جملات، هر جمله به خودرمزگذار ژرف ورودی داده می‌شود و بازنمایی سطح بالای آن بدست می‌آید که در این پژوهش با عبارت C(s) نمایش داده شده است. سپس، با استفاده از همین خودرمزگذار ژرف، بازنمایی پرس‌وجو بدست می‌آید که با عبارت C(q) نمایش داده می‌شود. سپس با استفاده از رابطه‌ی فاصله‌ی کسینوسی، میزان شباهت بازنمایی هر جمله‌ی s یعنی C(s)، با C(q) بدست می‌آید. میزان این شباهت‌ها، مبنای رتبه‌بندی جملات قرار می‌گیرد. با رتبه‌بندی جملات، خلاصه با انتخاب n جمله با بالاترین امتیاز تولید می‌شود. ورما و همکارانش مجموعه‌ای از ویژگی‌ها را برای انتخاب جملات با اهمیت استفاده نمودند [15]. سپس با استفاده از ماشین بولتزمن محدود، این ویژگی‌ها را به ویژگی‌های سطح بالاتری بازنمایی کردند که این پژوهش از آن با عنوان گسترش ویژگی نام برده است. سپس با استفاده از این ویژگی‌های سطح بالا، جملات با اهمیت برای تولید خلاصه انتخاب شده است. در پژوهشی دیگر، میرزاسلیمان و همکاران یک الگوریتم موازی با نام گریدی را به منظور خلاصه‌سازی خودکار متن در یک معماری توزیع‌شده ارائه نموده‌اند. این الگوریتم نیاز به حداقل ارتباطات را دارد و می‌توان آن را به سادگی با استفاده از مدل نگاشت-کاهش پیاده‌سازی نمود [16]. در تحقیق دیگری که توسط بادانیدیورو و همکارانش صورت پذیرفت، الگوریتمی برای خلاصه‌سازی روی داده‌های جویباری ارائه گردید که سیو-استریمینگ نام دارد و برای بیشینه‌سازی توابع زیرپیمانه‌ای یکنوا تحت محدودیت اندازه استفاده شده است. الگوریتم ارائه شده تنها به یک گذر با ترتیب دلخواه روی داده‌ها نیاز دارد و تضمین تخمین با ضریب ۱/۲-ε را نسبت به راه‌حل بهینه برای هر ε>0 ارائه می‌دهد [17].

٤- روش پیشنهادی

خلاصه‌سازی اخبار حول یک رویداد خاص را می‌توان به عنوان یک مساله‌ی خلاصه‌سازی چندسندی در نظر گرفت. در این مقاله، برای مدل‌سازی اخبار حول یک رویداد خاص، از گراف وزن‌دار $G = (V, E)$، که در آن هر راس مانند $v \in V$، نشان‌دهنده‌ی یک جمله موجود در اخبار، هر یال $e_{ij} \in E$، وجود ارتباط (شباهت) بین دو راس (جمله) $i, j \in V$ و وزن هر یال، تابعی مانند $w: E \to \mathbb{R}_+ \cup \{0\}$ است، استفاده می‌شود. این گراف، گراف زمینه‌ای سندها نامیده می‌شود.

مساله‌ی خلاصه‌سازی استخراجی چندسندی را به صورت رسمی زیر می‌توان بیان نمود: مجموعه‌ای از سندهای $D = \{d_1, d_2, \dots, d_m\}$ داده شده است که همگی اعضای D، سندهای متنی و دارای موضوع یکسانی هستند. هر کدام از سندها، مجموعه‌ای از جملات $d_i = \{s_1, s_2, \dots, s_{m_i}\}, (1 \le i \le m)$ است. در صورتی‌که $V = \cup_{i=1}^{m} \{s | s \in d_i\}$، مجموعه‌ی تمامی جملات موجود در سندهای متنی باشد، هدف از خلاصه‌سازی چندسندی استخراجی، تولید خلاصه‌ی $S^* = \{S_1, S_2, \dots, S_n\}$، با انتخاب جملات حاوی اطلاعات مفید و غیرافزونه از مجموعه‌ی V و با در نظر گرفتن محدودیت‌های C است. در صورتی‌که تابع هدف F، جملات را بر حسب میزان اطلاعات مفید و عدم افزونگی آنها ارزش‌گذاری نماید، در این صورت با بیشینه‌سازی F تحت محدودیت $\mathbb{C}$، خلاصه‌ی S به دست می‌آید. این موضوع در رابطه‌ی (۵) نمایش داده شده است.

$$S^* = argmax_{S \subseteq V}\{F(S)\} \; s.t. \; \sum_{s_i \in S} C(s_i) \le \mathbb{C}, 1 \le i \le n \qquad (۵)$$

در رابطه (۵)، تابع $C(.)$، مقدار هزینه‌ی جمله‌ی ورودی s_i را مشخص می‌کند. این تابع بایستی بر اساس تعریف محدودیت $\mathbb{C}$ مشخص گردد. به عنوان مثال، اگر محدودیت $\mathbb{C}$ حداکثر تعداد جملات تشکیل دهنده‌ی خلاصه

www.SID.ir

را مشخص نماید، می‌توان تابع هزینه را برای هر جمله، مقدار ثابت یک در نظر گرفت.

در این مقاله از یک تابع زیرپیمانه‌ای به عنوان تابع هدف استفاده گردیده است، زیرا امکان تضمین کیفیت خلاصه‌ی تولید شده را میسر می‌نماید. بنابراین، مساله خلاصه‌سازی چندسندی، معادل با مساله‌ی بیشینه‌سازی یک تابع زیرپیمانه F با محدودیت اندازه در نظر گرفته می‌شود. حال به منظور قابل یادگیری بودن پارامترهای تابع زیرپیمانه‌ای F، این تابع معادل با یک شبکه‌ی عصبی در نظر گرفته می‌شود. لایه‌ی ورودی این شبکه‌ی عصبی، ویژگی‌هایی هستند که در بخش استخراج ویژگی به آنها پرداخته می‌شود.

٤-١- استخراج ویژگی

ویژگی‌ها، لایه‌ی ورودی یک شبکه‌ی عصبی را تشکیل می‌دهند که معادل با تابع هدف F است. ویژگی‌های مورد استفاده به دو دسته‌ی کلی ویژگی‌های اهمیت و ویژگی‌های تاثیر تقسیم می‌شوند. ویژگی‌های اهمیت، ارزش هر جمله را مستقل از جملات دیگر در نظر می‌گیرند. برای مدل‌سازی این ویژگی‌ها، از توابع پیمانه‌ای استفاده می‌شود. ویژگی‌های تاثیر، ارزش هر جمله را با توجه به ارتباط آن با جملات دیگر در نظر می‌گیرند. ارتباط‌ها بین جملات از طریق گراف زمینه‌ای سندها، قابل دسترسی است. برای مدل‌سازی این دسته از ویژگی‌ها، از توابع زیرپیمانه‌ای استفاده شده است. ویژگی‌های اهمیت و تاثیر، به صورت توابع $f: 2^V \rightarrow \mathbb{R}_+ \cup \{0\}$ تعریف می‌شوند. از آنجایی که این ویژگی‌ها، لایه‌ی ورودی شبکه عصبی معادل با تابع هدف F را تشکیل می‌دهند، برای حفظ زیرپیمانگی این تابع، نیاز به بیان و اثبات قضایای ریاضی وجود دارد که در ادامه به آنها پرداخته می‌شود.

٤-١-١- ویژگی‌های اهمیت

ویژگی‌های اهمیت، ارزش هر جمله را مستقل از جملات دیگر در نظر می‌گیرند. برای مدل‌سازی این ویژگی‌ها، از توابع پیمانه‌ای استفاده می‌شود.

تعریف ۱: مجموعه‌ی واژگان کلیدی جمله s که با $K(s)$ نمایش داده می‌شود، عبارت است از آن دسته از واژه‌های جمله که با اجرای عملیات پیش‌پردازش حذف نمی‌شوند.

ویژگی‌های اهمیت شامل فراوانی واژگان، تعداد واژه‌های عنوان در متن(های) اصلی، موقعیت و نیز تعداد واژه‌های سیاقی در جمله است که خاصیت پیمانه‌ای بودن آنها در ادامه بررسی می‌شود.

تعریف ۲: تعداد وقوع واژه‌ی w در جمله‌ی s با $tf(w,s)$ نمایش داده شده و به آن، فراوانی واژه‌ی w در جمله‌ی w گفته می‌شود.

قضیه ۵ (ویژگی فراوانی واژگان): تابع $f: 2^V \rightarrow \mathbb{R}_+ \cup \{0\}$ که به ازای هر $A \subseteq V$، با استفاده از ضابطه‌ی $f(A) = \sum_{s \in A} \sum_{w \in K(s)} tf(w,s)$ ، قابل محاسبه است، پیمانه‌ای، نرمال و یکنوای غیر نزولی است.

اثبات: با توجه به ضابطه‌ی تابع f، واضح است که $f(\phi) = 0$، بنابراین، f، نرمال است.

اکنون تابع f در صورتی پیمانه‌ای است که به ازای هر $A \subseteq V$، داشته باشیم، $f(A) = \sum_{a \in A} f(a)$ فرض کنید که $A \subseteq V$، در این صورت به ازای هر $s \in A$، رابطه (۶) برقرار خواهد بود:

$$f(s) = \sum_{w \in K(s)} tf(w,s) \qquad (۶)$$

اکنون به ازای هر $A \subseteq V$، می‌توان نوشت:

$$f(A) = \sum_{s \in A} f(s) = \sum_{s \in A} \sum_{w \in K(s)} tf(w,s) \qquad (۷)$$

بنابراین، f پیمانه‌ای است.

از آنجایی که مقدار این تابع به ازای هر $s \in A$، نامنفی است، بنابراین با توجه به ضابطه‌ی آن، یکنوای غیر نزولی است.

قضیه ٦ (ویژگی تعداد واژگان عنوان): اگر T، نشان‌دهنده‌ی مجموعه‌ی واژگان عنوان یک سند متنی و sجمله‌ای از یک سند متنی باشد، در این صورت تابع $f: 2^V \rightarrow \mathbb{R}_+ \cup \{0\}$ که به ازای هر $A \subseteq V$، با استفاده از ضابطه‌ی $f(A) = \sum_{s \in A} |K(s) \cap T|$ قابل محاسبه است، پیمانه‌ای، نرمال و یکنوای غیر نزولی است.

اثبات: با توجه به ضابطه‌ی تابع f، واضح است که $f(\phi) = 0$، بنابراین، f، نرمال است.

اکنون تابع f در صورتی پیمانه‌ای است که به ازای هر $A \subseteq V$، داشته باشیم، $f(A) = \sum_{a \in A} f(a)$

فرض کنید که $A \subseteq V$، در این صورت به ازای هر $s \in A$، رابطه (۸) برقرار خواهد بود:

$$f(s) = |K(s) \cap T| \qquad (۸)$$

اکنون به ازای هر $A \subseteq V$ می‌توان نوشت:

$$f(A) = \sum_{s \in A} f(s) = \sum_{s \in A} |K(s) \cap T| \qquad (۹)$$

بنابراین، f پیمانه‌ای است.

از آنجایی که مقدار این تابع به ازای هر $s \in A$، نامنفی است، بنابراین با توجه به ضابطه‌ی آن، یکنوای غیر نزولی است.

از آنجایی که معمولا جملات آغازین متن از اهمیت بیشتری برخوردار هستند، می‌توان از ویژگی موقعیت جمله در متن به عنوان یکی از سنجه‌های ارزیابی اهمیت آن جمله استفاده نمود. مطابق قضیه ۷، ویژگی موقعیت جمله را نیز می‌توان به صورت یک تابع پیمانه‌ای در نظر گرفت.

قضیه ۷ (ویژگی موقعیت جمله):

تابع $f: 2^V \rightarrow \mathbb{R}_+ \cup \{0\}$ که به ازای هر $A \subseteq V$، با استفاده از ضابطه‌ی $f(A) = \sum_{s \in A} (1 - \frac{i(s)-1}{M-1})$ قابل محاسبه است، پیمانه‌ای، نرمال و یکنوای غیر نزولی است (i موقعیت جمله و M تعداد جملات سند متنی را نشان می‌دهد).

www.SID.ir

اثبات: با توجه به ضابطه‌ی تابع f، واضح است که $f(\phi)=0$، بنابراین، f، نرمال است.

اکنون تابع f در صورتی پیمانه‌ای است که به ازای هر $A \subseteq V$، داشته باشیم، $f(A)=\sum_{a\in A} f(a)$

فرض کنید که $A \subseteq V$، در این صورت به ازای هر $s \in A$، رابطه (۱۰) برقرار است:

$$f(s)=1-\frac{i(s)-1}{M-1} \quad (۱۰)$$

اکنون به ازای هر $A \subseteq V$ می‌توان نوشت:

$$f(A)=\sum_{s\in A} f(s)=\sum_{s\in A}(1-\frac{i(s)-1}{M-1}) \quad (۱۱)$$

بنابراین، f پیمانه‌ای است.

از آنجایی که مقدار این تابع به ازای هر $s \in A$، نامنفی است، بنابراین با توجه به ضابطه‌ی آن، یکنوای غیر نزولی است.

٤-۱-۲- ویژگی‌های تاثیر

ویژگی‌های تاثیر، ارزش هر جمله را با توجه به ارتباط آن با جملات دیگر در نظر می‌گیرند. این ارتباطها با کمک یک گراف وزن‌دار که گراف زمینه‌ای سند نامیده می‌شود، مدل‌سازی می‌گردد. برای مدل‌سازی این ویژگی‌ها، از توابع زیرپیمانه‌ای استفاده شده است.

قضیه ۸ (ویژگی پوشش راس وزن‌دار): اگر $G=(V,E)$، گراف زمینه‌ای یک سند، $A \subseteq V$ و $\delta(A)$ مجموعه‌ی همسایگان A در گراف G باشد، تابع $f: 2^V \to \mathbb{R}_+ \cup \{0\}$ با ضابطه‌ی $f(A)=\sum_{i\neq j,\ i,j\in A\cup\delta(A)} w_{ij}$، یک تابع زیرپیمانه‌ای، نرمال و یکنوای غیرنزولی است.

اثبات این قضیه در [3] آورده شده است.

با افزایش گوناگونی بین جملات، افزونگی کاهش می‌یابد. برای امتیازدهی به گوناگونی جملات، می‌توان گراف زمینه‌ای سندها، $G=(V,E)$ را بر اساس یک الگوریتم خوشه‌بندی به K خوشه‌ی P_i، $i=1,\dots,K$ افراز نمود. هر کدام از این خوشه‌ها یک جامعه‌ی افزونه را تشکیل می‌دهند. به عبارت دیگر، جملات مشابه (افزونه) در خوشه‌های یکسانی قرار می‌گیرند. در قضیه ۱۲، ویژگی گوناگونی بر اساس خوشه‌بندی مطرح شده، ارائه می‌گردد.

قضیه ۹ (ویژگی گوناگونی جملات): تابع $f: 2^V \to \mathbb{R}_+ \cup \{0\}$ به ازای هر $A \subseteq V$، با استفاده از ضابطه‌ی $f(A)=\sum_{i=1}^{K}\sqrt{\sum_{j\in P_i\cap A} r_j}$ قابل محاسبه است، زیرپیمانه‌ای، نرمال و یکنوای غیرنزولی است. در این رابطه، K، تعداد خوشه‌ها و P_i، $i=1,\dots,K$ خوشه‌ی i–ام است. همچنین، r_j نشان‌دهنده‌ی میزان اهمیت تکین هر کدام از جملات است و برای محاسبه‌ی آن، می‌توان از هر کدام از ویژگی‌های اهمیت استفاده نمود.

اثبات این قضیه در [5] آورده شده است.

٤-۲- تابع هدف زیرپیمانه‌ای

همانطور که قبلا گفته شد، به منظور تضمین کیفیت خلاصه‌ی تولید شده، از یک تابع زیرپیمانه‌ای به عنوان تابع هدف استفاده گردیده، و به منظور قابل یادگیری بودن پارامترهای آن، این تابع معادل با یک شبکه‌ی عصبی در نظر گرفته شده است. در بخش قبل، خصوصیات ویژگی‌های مورد استفاده در لایه ورودی شبکه عصبی مورد بررسی قرار گرفت. اکنون لازم است، قضایایی برای اثبات نرمال بودن، یکنوایی غیر نزولی و زیرپیمانگی تابع هدف F ارایه گردد. اثبات این قضایا برای ارایه‌ی تضمین‌های مطرح شده در قضیه ۱ لازم است.

قضیه ۱۰: تابع هدف F، به شرطی که وزن‌های شبکه عصبی معادل آن نامنفی باشند و توابع فعالیت موجود در آن توابع مقعری باشند، تابع یکنوای غیرنزولی و نرمال است.

اثبات: با توجه به اینکه بلوک‌های سازنده تابع هدف F، همان ویژگی‌های اهمیت و تاثیر هستند که در یک شبکه عصبی با وزن‌های نامنفی تحت توابع فعالیت مقعر با هم ترکیب می‌شوند، و تمامی این توابع یکنوای غیرنزولی و نرمال هستند، بنابر قضایای (۲) و (۳) یکنوایی غیرنزولی تابع هدف F حفظ می‌گردد. همچنین، بدیهی است که F نرمال است.

قضیه ۱۱: تابع هدف F، به شرطی که وزن‌های شبکه عصبی معادل آن نامنفی باشند و توابع فعالیت موجود در آن، توابع مقعری باشند، یک تابع زیرپیمانه‌ای است.

اثبات: با توجه به اینکه بلوک‌های سازنده تابع هدف F، توابع پیمانه‌ای و زیرپیمانه‌ای هستند که در یک شبکه عصبی با وزن‌های نامنفی تحت توابع فعالیت مقعر با هم ترکیب می‌شوند، بنابر قضایای (۲) و (۳) زیرپیمانه‌ای بودن تابع هدف F حفظ می‌گردد.

اکنون با توجه به اثبات‌های ارایه شده برای قضایای ۱۰ و ۱۱، این اطمینان حاصل می‌شود که تابع هدف F، تابعی نرمال، یکنوای غیرنزولی و زیرپیمانه‌ای است. بنابراین با استفاده از الگوریتم حریصانه امکان استخراج جملات خلاصه وجود دارد.

٤-۳- فرآیند یادگیری

به منظور یادگیری وزن‌های شبکه‌ی عصبی معادل تابع هدف F، نیاز به اجرای فرآیند یادگیری وجود دارد. با توجه به استفاده از پیکره‌های پاسخ و DUC 2004، و در دست بودن خلاصه‌ی مربوط به هر کدام از مجموعه‌های خبری حول یک موضوع، امکان ایجاد مجموعه‌ی آموزشی وجود دارد. برای برچسب‌گذاری داده‌های آموزشی، از تابع ROUGE-1 که در ادبیات خلاصه‌سازی متن شناخته شده است، استفاده می‌گردد. برای این منظور، مقدار ROUGE-1 بین سندهای خبری و خلاصه‌ی ارایه شده برای آنها در پیکره محاسبه می‌گردد. از طرف دیگر، برای تشکیل ورودی، مقدار ویژگی‌های اهمیت و تاثیر برای سندهای خبری محاسبه می‌شود. با استفاده از این اطلاعات، امکان آموزش شبکه‌های عصبی وجود دارد. با توجه به لزوم نامنفی

www.SID.ir

Arcl

بودن وزن‌ها، بسته‌های نرم‌افزاری مانند کراس امکاناتی را بـرای ایـن منظـور ارایه می‌دهند. به این ترتیب ایـن اطمینـان حاصـل مـی‌شـود کـه وزن‌هـای یادگیری شده نامنفی هستند و خللـی در تضـمین زیرپیمـانگی کـل شـبکه‌ی عصبی وارد نمی‌نمایند.

۵- نتیجه

به منظور ارزیابی این مقاله، از پیکره پاسخ برای زبان فارسی و پیکره DUC 2004 برای زبان انگلیسی استفاده شده است. این پیکره‌ها شامل اخبار حـول موضوعات اجتماعی، سیاسی، ورزشی و غیره اسـت. پیکـره پاسـخ حـاوی ۵۰ موضوع است که در مورد هر کدام از موضوعات، ۲۰ سند خبـری وجـود دارد. خلاصه‌های مرجع به روش استخراجی توسط عامل‌های انسانی برای هر کدام از موضوعات تولید شده است. در طرف دیگر، پیکره DUC 2004 شامل ۵۰ موضوع است که برای هر کدام از موضوعات، ۱۰ سند خبری ارایه شده است. جدول (۱)، نتایج خلاصه‌ساز پیشـنهادی بـر مبنـای پیکـره پاسـخ را نمـایش می‌دهد. متاسفانه به دلیل عدم وجود نتایج مشابه از پژوهش‌های دیگر در زبان فارسی، امکان مقایسه بین روش‌ها میسر نبوده است.

در جدول (۲)، به مقایسه روش پیشنهادی با روش‌های جدید خلاصه‌سـازی ارایه شده برای زبان انگلیسی پرداختـه شـده، کـه بـر مبنـای پیکـره DUC 2004، صورت پذیرفته است.

همانطور که در جدول (۲) مشخص است، عملکرد روش پیشنهادی نسبت بـه روش‌های دیگـر بـر اسـاس معیارهـای ROUGE-1 و ROUGE-2 بهتـر است و امتیاز ROUGE-L بدست آمده، با تنها مقاله‌ای که این امتیـاز در آن گزارش گردیده، قابل مقایسه است.

جدول (۱): ارزیابی روش با استفاده از پیکره پاسخ

ROUGE-1	ROUGE-2	ROUGE-L
۳۸/۵	۱۲/۶	۳۱/۴

جدول (۲): ارزیابی روش با استفاده از پیکره DUC 2004

	ROUGE-1	ROUGE-2	ROUGE-L
روش پیشنهادی	۴۰/۲۸	۱۰/۲۳	۲۵/۸۱
ICSISumm [18]	۳۸/۴۷	۹/۷۸	N/A
RegSum [19]	۳۸/۵۷	۹/۷۵	N/A
DPP [20]	۳۹/۷۹	۹/۶۲	N/A
R2N2-ILP [11]	۳۸/۷۸	۹/۸۶	N/A
GRU+GCN [21]	۳۸/۲۳	۹/۴۸	N/A
Uckan [22]	۳۸/۰۷	۸/۲۷	۲۹/۰۳

مراجع

[1] J. M. Torres-Moreno, Automatic Text Summarization, Wiley, 2014.

[2] A. Krause , D. Golovin, "Submodular function maximization *Tractability: Practical Approaches to Hard Problems* , Cambridge University Press, 2013.

[3] L. Lovász, "Submodular functions and convexity ", *Mathematical Programming The State of the Art* , Springer, 1983 ,pp. 235-257.

[4] H. Lin , J. Bilmes, "Multi-document summarization via budgeted maximization of submodular functions ",in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* ,2010.

[5] H. Lin , J. Bilmes, "A class of submodular functions for document summarization ",*Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1* , 2011.

[6] G. L. Nemhauser, L. A. Wolsey , M. L. Fisher, "An analysis of approximations for maximizing submodular set functions—I ",*Mathematical Programming* ,14 ,pp. 265-294, 1978.

[7] B. Mirzasoleiman, A. Karbasi, R. Sarkar , A. Krause, "Distributed submodular maximization: Identifying representative elements in massive data در ",*Advances in Neural Information Processing Systems* ,2013 .

[8] K. Wei, Y. Liu, K. Kirchhoff , J. Bilmes, "Unsupervised submodular subset selection for speech data ",*Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on* ,2014 .

[9] J. Kupiec, J. Pedersen , F. Chen, "A trainable document summarizer ",*Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval* , 1995 .

[10] B. Larsen, "A trainable summarizer with knowledge acquired from robust NLP techniques ",*Advances in automatic text summarization* ,71 ,1999 .

[11] Z. Cao, F. Wei, L. Dong, S. Li , M. Zhou, "Ranking with Recursive Neural Networks and Its Application to Multi-Document Summarization در ",*AAAI* ,2015 .

[12] P.-y. Zhang , C.-h. Li, "Automatic text summarization based on sentences clustering and extraction ", *Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on* , 2009 .

[13] R. A. García-Hernández , Y. Ledeneva, "Word sequence models for single text summarization ", *Advances in Computer-Human Interactions, 2009. ACHI'09. Second International Conferences on* ,2009 .

[14] M. Yousefi-Azar و L. Hamey, "Text Summarization Using Unsupervised Deep Learning ",*Expert Syst. Appl* ,.68 ,pp. 93-105, 2 2017 .

[15] P. Verma , H. Om, "MCRMR: Maximum coverage and relevancy with minimal redundancy based multi-document summarization ",*Expert Systems with Applications* ,120 ,pp. 43-56, 2019 .

[16] B. Mirzasoleiman, A. Karbasi, R. Sarkar , A. Krause, "Distributed Submodular Maximization: Identifying

www.SID.ir

Arcl

Representative Elements in Massive Data ", *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2* , Red Hook, NY, USA, 2013 .

[17] A. Badanidiyuru, B. Mirzasoleiman, A. Karbasi و ,A. Krause, "Streaming submodular maximization: Massive data summarization on the fly ",*Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* ,2014 .

[18] D. Gillick, B. Favre, D. Hakkani-Tür, B. Bohnet, Y. Liu, S. Xie, "The ICSI/UTD Summarization System at TAC 2009 ",.*TAC* ,2009 .

[19] K. Hong , A. Nenkova, "Improving the Estimation of Word Importance for News Multi-Document Summarization", *Proceedings of the 1٣th Conference of the European Chapter of the Association for Computational Linguistics* ,Gothenburg, 2014 .

[20] A. Kulesza , B. Taskar, Determinantal Point Processes for Machine Learning, Hanover, MA :Now Publishers Inc., 2012 .

[21] M. Yasunaga, R. Zhang, K. Meelu, A. Pareek, K. Srinivasan , D. Radev, "Graph-based Neural Multi-Document Summarization ",*Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)* , Vancouver, 2017 .

[22] T. Uçkan , A. Karcı, "Extractive multi-document text summarization based on graph independent sets ", *Egyptian Informatics Journal* ,2020 .

www.SID.ir