مقایسهٔ روش‌های مختلف یادگیری ماشین در خلاصه‌سازی استخراجی گفتار به گفتار فارسی بدون استفاده از رونوشت

هدی سادات جعفری* **و محمدمهدی همایون‌پور**

آزمایشگاه پردازش هوشمند داده‌های چندرسانه‌ای، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، ایران

چکیده

در این مقاله، خلاصه‌سازی استخراجی گفتار با استفاده از روش‌های مختلف یادگیری ماشین مورد مطالعه قرار گرفته است. خلاصه‌سازی یک فایل گفتاری به معنای استخراج بخش‌های مهم و شاخص گفتار به‌منظور دسترسی، جستجو، استخراج و مرورگری آسان‌تر و کم‌هزینه‌تر اطلاعات فایل‌های گفتاری است. در این مقاله، یک روش جدید خلاصه‌سازی گفتار بدون استفاده از سامانهٔ بازشناسی خودکار گفتار ارائه شده است. الگوهای تکراری بین دو جملهٔ گفتاری با استفاده از الگوریتم S-DTW، به‌طورمستقیم از روی سیگنال گفتار شناسایی می‌شوند. بعد از تعیین شباهت بین دو جمله و استخراج تعدادی ویژگی از هر جمله تأثیر روش‌های مختلف یادگیری ماشین، بانظارت، بی‌نظارت و نیمه‌نظارتی مورد بررسی قرار گرفته است. آزمایش‌ها برروی یک پیکرهٔ خوانده‌شدهٔ اخبار فارسی انجام شده است. نتایج نشان می‌دهد با استفاده از ویژگی‌های مناسب، بدون استفاده از رونوشت به کارایی بالاتری نسبت به روش‌های پایه (۳٪ افزایش در مقایسه با انتخاب نخستین جملات و ۵٪ افزایش در مفایسه با انتخاب طولانی‌ترین جملات با استفاده از معیار ROUGE-3) می‌توان دست پیدا کرد.

واژگان کلیدی: خلاصه‌سازی استخراجی گفتار، سیگنال گفتار، الگوها کلیدی، الگوریتم S-DTW، یادگیری ماشین

A comparison of machine learning techniques for Persian Extractive Speech to Speech Summarization without Transcript

Hoda Sadat Jafari *& Mohammad Mehdi Homayounpour

Laboratory for Intelligent Multimedia Processing, Amirkabir University of Technology, Tehran, Iran

Abstract

In this paper, extractive speech summarization using different machine learning algorithms was investigated. The task of Speech summarization deals with extracting important and salient segments from speech in order to access, search, extract and browse speech files easier and in a less costly manner. In this paper, a new method for speech summarization without using automatic speech recognition system (ASR) is proposed. ASR systems usually have high error rates especially in adverse acoustic environment and for low resource languages. Our goal was to answer this question: is it possible to summarize a Persian speech without ASR using less or no training data? We have proposed a method which discovers salient parts directly from speech signal by using a semi-supervised algorithm. The proposed algorithm consists of three main stages, features extraction, identifying key patterns and selecting important sentences.

First we have segmented speech voices manually into sentences to eliminate sentence segmentation errors. Therefore, we could have better comparison between different summarization methods. Then we have extracted some features from each sentence such as sentence duration, if the sentence is first or last sentence in the speech and so on. Also, repetitive patterns between each two sentence of speech are discovered directly from speech signal by using S-DTW algorithm. S-DTW algorithm can discover repetitive patterns between

* Corresponding author

* نویسندهٔ عهده‌دار مکاتبات

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.143]

two speech signals by using MFCC features. By using these repetitive patterns between each pair of sentences we can make a similarity matrix. Therefore, we could measure the similarity distance between each pair of sentences and eliminate redundant sentences from summary without the need to use an ASR system

After finding the similarity between each two speech segments and extracting some features from each segment, various machine learning algorithms including unsupervised (MMR, TextRank), supervised (SVM, Naïve Bayes) and semi-supervised algorithms (self-training, Co-training) are used in order to extract salient parts. Experiences are done in read Persian news. The results show that using semi-supervised co-training method and appropriate features, the performance of speech summarization system on read Persian news corpus can improve about 3% compared to selecting the first sentences and by 5% compared to longest sentences when ROUGE-3 is used as the evaluation measure.

Key words: Extractive speech summarization, speech signal, key patterns, S-DTW algorithm, machine learning

۱- مقدمه

در این مقاله، خلاصه‌سازی استخراجی گفتار با استفاده از روش‌های مختلف یادگیری ماشین مورد مطالعه قرار گرفته است. خلاصه‌سازی یک فایل گفتاری به معنای استخراج بخش‌های مهم و شاخص گفتار به‌منظور دسترسی، جستجو، استخراج و مرورگری آسان‌تر و کم‌هزینه‌تر اطلاعات فایل‌های صوتی است. امروزه، با رشد روزافزون اطلاعات و قرارگرفتن حجم بالایی از فایل‌های صوتی برروی شبکه‌ها و رایانه‌ها، یافتن راه‌حل‌هایی کارا برای ذخیره‌سازی، بازیابی، جستجو و مرورگری سریع این فایل‌های صوتی مورد نیاز است. خلاصه‌سازی گفتار یکی از راه‌های پیشنهادی برای حل مشکلات بالا است.

سامانهٔ خلاصه‌سازی گفتار[1] در زمینه‌های مختلف، خلاصه‌سازی جلسات، کنفرانس‌ها، کلاس‌های درس، اخبار رادیو و تلویزیون و ... مورد استفاده می‌تواند قرار بگیرد. خروجی سامانهٔ خلاصه‌سازی گفتار در مرورگرها، موتورهای جستجو، سامانه‌های بازیابی و بازیابی اطلاعات[2] و نشانه‌گذاری[3] کاربرد دارد [1]، [2]. در این مقاله خلاصه‌سازی گفتار خوانده‌شده از روی متون خبری فارسی مورد بررسی قرار گرفته است.

در خلاصه‌سازی گفتار مشابه خلاصه‌سازی متن با چالش‌هایی همچون انتخاب ویژگی و روش انتخاب بهترین جملات برای خلاصه مواجه هستیم؛ علاوه‌بر آن خلاصه‌سازی گفتار مشکلات خاصی دارد که در ادامه به برخی از آنها اشاره می‌شود.

سامانه‌های مرسوم خلاصه‌سازی گفتار به‌طورمعمول شامل دو بخش اصلی تبدیل گفتار به متن با استفاده از یک سامانه بازشناسی خودکار گفتار[4] و سپس خلاصه‌سازی متن تولیدشده، می‌شوند. مشکل اصلی در این خلاصه‌سازها انواع خطاهای ناشی از بازشناسی خودکار گفتار است. سامانهٔ بازشناسی خودکار گفتار، در شرایط صوتی سخت[5] (مانند محیط‌های نوفه‌ای، گفتار تلفنی و گفتار حاوی انعکاس محیط و مانند آن) و وقتی زبان دارای محدودیت داده‌های آموزشی است، کارایی پایینی دارد. دقت پایین این سامانه‌های خلاصه‌سازی گفتار باعث به‌وجودآمدن اطلاعات غلط و تشخیص نادرست کلمات در اثر خطاهای بازشناسی می‌شود [2]، [3]. مشکلات بالا باعث کاهش دقت در بازشناسی خودکار گفتار و درنتیجه باعث کاهش دقت در خلاصه‌سازی گفتار می‌شود. علاوه بر آن، در تمام سامانه‌های بازشناسی، دقت رونوشت[6] به میزان و کیفیت داده‌های آموزشی که برای مدل‌سازی آکوستیکی و مدل‌سازی زبانی استفاده شده است، بستگی دارد. تهیه داده‌های آموزشی بسیار زمان‌بر و هزینه‌بر است و برای بسیاری از زبان‌ها چنین داده‌هایی وجود ندارد. همچنین، چون سامانهٔ بازشناسی خودکار گفتار از یک لغت‌نامه[7] از قبل مشخص‌شده استفاده می‌کنند، ممکن است لغاتی در داده‌های آزمون وجود داشته باشد که این لغات خارج از لغت‌نامه[8] (OOV) باشند. نرخ OOVها با افزایش حجم لغت‌نامه می‌تواند کاهش پیدا کند؛ اما، بزرگ‌شدن بیش از حد لغت‌نامه و داشتن لغات فرعی[9] زیاد نیز مناسب نیست.

[1] Speech summarization system
[2] Information retrieval
[3] Indexing

[4] Automatic speech recognition
[5] Adverse acoustic
[6] Transcript
[7] Vocabulary
[8] Out Of Vocabulary
[9] Extraneous

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.143]

در متن به‌طورمعمول نظم خاصی بین جملات و بندها برقرار است. تشخیص جملات در متن با استفاده از علائمی که در متن به کار می‌رود کار آسانی است؛ درحالی‌که در گفتار تشخیص جملات کار دشواری است. بنابراین نبود چنین علائمی در رونوشت تولیدشده از سامانهٔ بازشناسی خودکار گفتار باعث تشخیص نادقیق بندها و جملات می‌شود [2]. روانی[1] گفتار در مواردی چون شروع مجدد[2]، بازنگری در کلام[3]، تکرارها[4]، ایجاد صداهای غیر گفتاری[5] از دست می‌رود که این ویژگی در گفتار، کار سامانه بازشناسی خودکار گفتار را مشکل‌تر می‌کند [2]،[3]. موارد بالا خلاصه‌سازی گفتار را دچار چالش‌های جدیدی نسبت به خلاصه‌سازی متن می‌کند.

در خلاصه‌سازی فایل‌های گفتاری داشتن چندین گوینده، بحث‌های خارج از موضوع[6]، گفتگوهای کوتاه[7]، جملات محاوره‌ای و عدم رعایت گرامر زبان، هم‌پوشانی بین گفتارها از دیگر مشکلات این بخش هستند. همچنین کیفیت پایین تجهیزات ضبط باعث افزایش خطای بازشناسی خودکار گفتار می‌تواند شود [2].

برای این‌که سامانهٔ خلاصه‌سازی گفتار بتواند مستقل از سامانهٔ بازشناسی خودکار گفتار باشد و از برخی چالش‌های بازشناسی خودکار گفتار صرف نظر کنیم، در این مقاله روشی برای خلاصه‌سازی فایل‌های گفتاری بدون استفاده از سامانهٔ بازشناسی خودکار گفتار ارائه شده است. در این مقاله، خلاصه‌سازی استخراجی، تک‌سنده، تک‌گوینده، ورودی و خروجی گفتار فارسی مورد مطالعه قرار گرفته است. روش پیشنهادی مستقل از سامانهٔ بازشناسی خودکار گفتار بوده و خلاصه به‌طورمستقیم از روی سیگنال گفتار استخراج می‌شود. تأثیر روش‌های مختلف یادگیری ماشین، بانظارت، بی‌نظارت و نیمه‌نظارتی مورد بررسی قرار گرفته است. درواقع هدف از این مقاله بررسی امکان تولید خلاصه‌ای از روی سیگنال گفتار بدون بازشناسی گفتار با استفاده از روش‌های یادگیری ماشین است. تعدادی از روش‌های یادگیری ماشین که در پژوهش‌های گذشته کارایی بهتری در این زمینه داشته‌اند، انتخاب و مورد آزمایش قرار گرفتند. رویکرد پیشنهادی نیاز به استفاده از سامانه‌های بازشناسی گفتار را که به‌طورمعمول به‌راحتی و به‌ارزانی در اختیار نیستند، مرتفع و از پیچیدگی‌های خلاصه‌سازی گفتار می‌کاهد.

در بخش دوم به بررسی کارهای مرتبط در زمینهٔ خلاصه‌سازی گفتار می‌پردازیم. روش پیشنهادی را در بخش سوم ارائه می‌کنیم؛ سپس، در بخش چهارم نتایج آزمایش‌های روش پیشنهادی را مورد بررسی قرار می‌دهیم و درنهایت در بخش پنجم نتیجه‌گیری و کارهای آینده ارائه می‌شود.

۲- کارهای مرتبط

دو رویکرد کلی برای خلاصه‌سازی گفتار وجود دارد: ۱) خلاصه‌سازی با استفاده از سامانهٔ بازشناسی خودکار گفتار ۲) خلاصه‌سازی بدون استفاده از سامانهٔ بازشناسی خودکار گفتار.

در روش نخست، ابتدا سیگنال گفتار با استفاده از سامانهٔ بازشناسی خودکار گفتار به متن تبدیل می‌شود؛ سپس، با استفاده از روش‌های رایجی که در خلاصه‌سازی متن وجود دارد، خلاصه‌ای از متن تولیدشده، ایجاد می‌شود. از معایب این روش خطای زیاد سامانهٔ بازشناسی خودکار گفتار به‌خصوص در گفتار محاوره‌ای و زمانی که کلمات خارج از واژگان زیادی وجود داشته باشد، است. خطاهای بازشناسی تأثیرات منفی‌ای برروی خلاصه‌سازی گفتار به‌خصوص در حوزه جلسات که به‌طورمعمول نرخ خطای کلمه در آن به‌دلیل محاوره‌ای‌بودن گفتار بیشتر از سایر حوزه‌هاست، می‌گذارد. به همین دلیل، در [4] از بهترین فرضیه‌ها برای بهبود بازشناسی استفاده شده است. در [5] علاوه‌بر بهترین فرضیه‌ها از شبکه ابهام[8] نیز برای بازشناسی بهتر جهت بهبود کارایی در خلاصه‌سازی جلسات ارائه شده است. در این دو مقاله، از روش بیشینه ارتباط حاشیه‌ای[9] (MMR) برای خلاصه‌سازی استفاده شده است. این الگوریتم به‌صورت تکراری جملاتی را برای خلاصه انتخاب می‌کند که بیشترین ارتباط را با جملات درون متن و کمترین افزونگی را با جملات خلاصه داشته باشند. در [6] و [7] از ویژگی‌ها آکوستیکی[10]/نوایی[11] به همراه ویژگی‌های لغوی و

[1] Fluency
[2] Restart
[3] Revisions
[4] Repetitions
[5] Filler words
[6] Off topic discussions
[7] Chit chat

[8] Confusion network
[9] Maximal Marginal Relevance
[10] Acoustic
[11] Prosodic

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]
[DOI: 10.29252/jsdp.14.4.143]

دسته‌بندی‌کنندهٔ ماشین بردار پشتیبان[1] (SVM) برای خلاصه‌سازی گفتار اخبار و سخنرانی استفاده شده است. در [8] از مدل SVM برای خلاصه‌سازی استخراجی جلسات با استفاده از ویژگی‌های مختلف لغوی، ساختاری، گفتمان و ویژگی‌های مرتبط با موضوع استفاده شده است. تأثیر هر یک از ویژگی‌ها با استفاده از الگوریتم انتخاب ویژگی رو به جلو[2] (FFS) بررسی شده و در [9] نیز الگوریتم SVM و برای بهبود دسته‌بندی از سه روش نمونه‌برداریِ بالا[3]، نمونه‌برداریِ پایین[4] و نمونه‌برداریِ دوباره[5] استفاده شده است. در [10] و [11] چارچوب کاهش ریسک[6] برای خلاصه‌سازی گفتار استفاده شده است. در [12]، اطلاعات بلاغی با استفاده از HMM مدل شده، برای خلاصه‌سازی گفتار استفاده شده است. در [13] نیز پژوهش‌های مشابهی انجام شده که در آن به جای استفاده از HMM از SVM استفاده شده و در [14] برای تعیین میزان اهمیت جملات از معیار واگرایی کولبک-لیبلر[7] استفاده شده است. در [15] نیز بعد از تعیین تعدادی ویژگی لغوی و نوایی از الگوریتم آموزش هم‌آموز[8] برای آموزش دسته‌بندی‌کننده استفاده شده که در این الگوریتم فرض شده است ویژگی‌ها به دو دسته مستقل (لغوی و نوایی) از هم می‌توانند تقسیم شوند و هر دسته برای آموزش یک دسته‌بندی خوب کفایت می‌کند. در [16] نیز از مدل‌های زبانی به همراه معیار واگرایی کولبک-لیبلر برای انتخاب جملات مهم استفاده شده و آزمایش‌ها برروی اخبار به زبان چینی انجام گرفته است. در [17] نیز روشی برای مرتب‌کردن جملات براساس اهمیت برای قرار گرفتن در خلاصه با استفاده از SVM معرفی کرده است.

در [18] و [19] نیز از الگوریتم MMR استفاده شده است. حالت حریصانهٔ الگوریتم MMR باعث می‌شود تا بهترین زیرمجموعه از جملات برای خلاصه پیدا نشود. بنابراین در این مقالات سعی شده است با تغییر تابع MMR و تعریف یک تابع هدف سراسری و حل یک مسئله بهینه‌سازی این مشکل حل شود. در [20] روشی با استفاده از سیر تصادفی[9] بر روی گرافی که با استفاده از عبارت‌های کلیدی استخراج‌شده به‌صورت خودکار ساخته شده و تحلیل معنایی پنهان احتمالاتی[10] (PLSA) ارائه شده است. در [21] و [22] نیز از مدل PLSA برای استخراج موضوع و خلاصه‌سازی استفاده شده است. در [23] از مدل‌های زبانی به همراه شبکه عصبی در خلاصه‌سازی اخبار رادیو و تلویزیون استفاده شده است. در [24] خلاصه‌سازی جلسات با استفاده از تقسیم‌بندی رونوشت گفتار به بخش‌هایی با توجه به گوینده غالب و یا گویندگانی که در هر بحث شرکت داشته‌اند، مورد بررسی قرار گرفته است. جملات با توجه به اینکه در کدام بخش گفتاری قرار گرفته‌اند، امتیازدهی می‌شوند. در [25] و [26] به جای بازشناسی کامل گفتار تنها از بازشناسی واج استفاده شده است. سعی شده تا خلاصه‌سازی و تشخیص موضوع تنها با استفاده از اطلاعات واجی و بدون بازشناسی کامل انجام شود. همان‌طورکه گفته شد، روش‌های دستهٔ نخست وابسته به سامانهٔ بازشناسی خودکار گفتار هستند و خطاهای بازشناسی خودکار گفتار از مهم‌ترین چالش‌های این دسته از روش‌هاست.

در روش دوم، سعی می‌شود بدون استفاده از سامانهٔ بازشناسی خودکار گفتار، شاخص‌بودن جملات به‌طورمستقیم از روی سیگنال گفتار به‌دست آورده شود. این روش‌ها نسبت به روش‌های دسته نخست مشکل‌تر هستند؛ اما به‌دلیل پتانسیلی که برای پردازش گفتار بدون استفاده از بازشناسی خودکار گفتار دارند، در سال‌های اخیر در برنامه‌های کاربردی مانند خلاصه‌سازی، خوشه‌بندی موضوعی و ... مورد توجه قرار گرفته‌اند [27] ، [28]، [29].

مقالهٔ [30] از نخستین کارهای انجام‌شده برای خلاصه‌سازی گفتار تنها با استفاده از ویژگی‌های نوایی و بدون استفاده از ویژگی‌های لغوی است. در این مقاله، از مدل HMM استفاده شده است. نتایج مبین استفاده از ویژگی‌های نوایی باعث بهبود سامانهٔ خلاصه‌ساز شده است. در [31] به هر هجا امتیازی با توجه طول هجا، تغییرات انرژی و گام داده می‌شود؛ سپس هر بخش از گفتار با توجه به امتیاز هجاهایش امتیازدهی می‌شود. بخش‌ها با توجه به امتیازشان در خلاصه قرار می‌گیرند. در [32] از گراف دولایه برای خلاصه‌سازی بی‌نظارت گفتار تنها با استفاده از ویژگی‌های نوایی و بدون داشتن اطلاعات لغوی استفاده و در [33] یک

[1] Support Vector Machine
[2] Forward Feature Selection
[3] Up-sampling
[4] Down-sampling
[5] Re-sampling
[6] A risk-aware modeling framework
[7] Kullback-Leibler (KL) divergence measure
[8] Co-training algorithm
[9] Random walk
[10] Probabilist Latent Semantic Analysis

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]
[DOI: 10.29252/jsdp.14.4.143]

روش خلاصه‌سازی گفتار برای سخنرانی به زبان چینی ارائه شده است. تأثیر ویژگی‌های مختلف آکوستیکی، لغوی و ساختاری مورد بررسی قرار گرفته است. از SVM با کرنل RBF برای دسته‌بندی جملات استفاده و نشان داده شده است که تنها با استفاده از ترکیبی از ویژگی‌های آکوستیکی و ساختاری که مستقل از ویژگی‌های لغوی هستند، به دقت خوبی در خلاصه‌ساز دست می‌توان پیدا کرد. در مقاله [23] روش تعبیه کلمه و پاراگراف برای خلاصه‌سازی متون گفتاری با یکدیگر مقایسه شده و همچنین در روش پیشنهادی شاخص ارتباط و افزونگی در جملات را به‌طور همزمان در نظر گرفته است. درنهایت از تعبیه پاراگراف برای بهبود کارایی خلاصه‌ساز استفاده شده است. یکی دیگر از روش‌هایی که در سال‌های اخیر مورد توجه قرار گرفته، استفاده از یادگیری عمیق است. در مقاله [9] از این روش برای خلاصه‌سازی استخراجی گفتار اخبار رادیو و تلویزیون به زبان چینی استفاده شده و نتایج نشان داده است که با استفاده از ترکیبی از ویژگی‌های لغوی و نوایی به کارایی خوبی در این زمینه می‌توان دست پیدا کرد.

از معایب روش‌هایی که از ویژگی‌های نوایی استفاده می‌کنند، این است که ویژگی‌های نوایی در همه حوزه‌ها نمی‌توانند به خلاصه‌سازی کمک کنند و بیشتر در حوزهٔ خلاصه‌سازی اخبار که در آن گویندگان آموزش‌دیده هستند، کاربرد دارند. همچنین استفاده از این ویژگی‌ها ما را به سمت افزونگی کمتر در خلاصه نمی‌تواند هدایت کند؛ بنابراین، استفاده از این ویژگی‌ها در هنگام خلاصه‌سازی چندسنده گفتار مناسب نیست [34]. درنتیجه در ادامه روش‌های دیگری برای حل مشکلات بالا معرفی شدند. در این روش‌ها سعی می‌شود تا با تغییر الگوریتم DTW[1]، الگوهای تکراری در یک سیگنال گفتار پیدا شود. در [35] یک روش بی‌نظارت کشف الگو در گفتار با نام الگوریتم پیچش زمانی پویای قطعه‌ای[2] (S-DTW) ارائه شده است. در الگوریتم DTW، فاصله اکوستیکی دو بردار محاسبه می‌شود؛ اما از آن برای پیداکردن زیردنباله‌های مشترک در دو دنباله بزرگ‌تر نمی‌توان استفاده کرد؛ زیرا نقطه شروع و پایان الگوها در چنین حالتی مشخص نیست؛ اما، در الگوریتم S-DTW با تعریف محدودیت‌هایی برروی الگوریتم اصلی DTW، زیردنباله‌های مشترک در دو بخش از گفتار را می‌توان پیدا کرد. در مقالات مختلف نیز روش S-DTW بهبود داده شده است تا بتواند در شرایط مختلف، به‌عنوان‌مثال با وجود چندین گوینده، و با هزینه محاسباتی کمتر الگوهای تکراری بین دو بخش از گفتار را پیدا کند. از این روش برای کشف، تشخیص، جستجوی عبارات گفتاری[3] [36]، [37] و [38] و خلاصه‌سازی گفتار [34]، [29] و [39] استفاده شده است. مزیت این روش نسبت به روش‌هایی که از ویژگی‌های نوایی استفاده می‌کنند، این است که علاوه‌بر اینکه احتیاج به هیچگونه داده آموزشی ندارد، اطلاعات بیشتری نسبت به ویژگی‌های نوایی به ما می‌دهد. در این مقاله نیز از الگوریتم S-DTW برای تعیین الگوهای تکراری بین هر دو جمله از گفتار برای تعیین شباهت بین دو جملهٔ گفتاری استفاده شده است.

۳- روش پیشنهادی

در روش پیشنهادی، ابتدا سیگنال ورودی به جملات تشکیل‌دهنده‌اش تقسیم می‌شود. بعد از فرم‌بندی جملات گفتاری و استخراج ویژگی‌های MFCC از هر فریم، از الگوریتم S-DTW [35] برای استخراج الگوهای تکراری بین هر دو جمله گفتاری استفاده می‌شود. در الگوریتم S-DTW مشابه الگوریتم DTW، بردارهای ویژگی دو سیگنال گفتار به‌صورت غیرخطی به یکدیگر متناظر شده و فاصله آکوستیکی بین دو بردار ویژگی محاسبه می‌شود. با این تفاوت که در الگوریتم S-DTW پیشنهاد شده است با درنظرگرفتن نقاط شروع و پایان مختلف در فضای جستجوی DTW، چندین مسیر ترازبندی بین دو رشته پیدا شود و بدین ترتیب از آن برای تعیین الگوهای تکراری بین دو جمله گفتاری استفاده شود. با استفاده از الگوهای تکراری پیداشده ماتریس شباهت بین هر دو جمله ساخته می‌شود. ماتریس شباهت یک ماتریس مربعی متقارن به تعداد جملات متن است که میزان شباهت بین هر دو جمله را مشخص می‌کند. بعد از تعیین شباهت بین هر دو جمله گفتاری، از سه دسته یادگیری بی‌نظارت، بانظارت و نیمه‌نظارتی برای تعیین جملات خلاصه استفاده شده است. روش‌های استفاده در هر بخش در ادامه توضیح داده می‌شود.

۳-۱- خلاصه‌سازی بی‌نظارت

در روش یادگیری بی‌نظارت، هدف پیداکردن یک قاعده و

[1] Dynamic Time Warping
[2] Segmental Dynamic Time Warping
[3] Spoken term detection

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]
[DOI: 10.29252/jsdp.14.4.143]

نظم در داده‌های ورودی است. بنابراین هدف در خلاصه‌سازی بی‌نظارت این است که جملات متن (داده‌های ورودی) را به دو دستهٔ خلاصه و غیرخلاصه تقسیم کنیم، بدون آنکه در داده‌های آموزشی مشخص شده باشد کدام جملات خلاصه و کدام جملات غیرخلاصه هستند. بدین منظور از دو روش MMR و TextRank استفاده شده است.

۱-۱-۳- روش بیشینهٔ ارتباط حاشیه‌ای (MMR)

این روش، یک روش حریصانه و بی‌نظارت است. برای هر جمله S_i در متن D، امتیاز $MMR(S_i)$ به‌صورت فرمول (۱) محاسبه می‌شود. در این فرمول $Sim(S_i, D)$ میزان شباهت این جمله را با کل متن و $Sim(S_i, Summ)$ میزان شباهت این جمله را با خلاصه‌ای که تا این لحظه تولید شده است، نشان می‌دهد. Summ نشان‌دهندهٔ خلاصهٔ تولیدشده تاکنون است. β فاکتور تعادل است. در پیاده‌سازی β برابر ۰/۵ قرار داده شده است.

$$MMR(S_i) = \beta \times Sim(S_i, D) - (1-\beta) \times Sim(S_i, Summ) \quad (۱)$$

در هر مرحله از تکرار الگوریتم، جمله‌ای که بیشترین امتیاز MMR را گرفته باشد، به‌صورت حریصانه انتخاب شده و در خلاصه قرار می‌گیرد. درواقع در این روش برای تولید خلاصه، ابتدا الگوهای تکراری در متن (کلمات) پیدا شده و سپس براساس این الگوها میزان شباهت جملات محاسبه شده و درنهایت یک خلاصه که در برگیرندهٔ اطلاعات متنوعی از متن اصلی با افزونگی کمی است، تولید می‌شود. برای استفاده از روش MMR در گفتار از ماتریس شباهتی که با استفاده از الگوریتم S-DTW ساخته می‌شود، استفاده می‌شود. ابتدا، امتیاز شباهت بین هر دو جمله با توجه به فرمول (۲) محاسبه می‌شود:

(۲)

$$sim(S_i, S_j) = \frac{similarityMatrix(S_i, S_j) \times pattern-length}{length(S_i) + length(S_j)}$$

پارامتر Sim امتیاز شباهت بین جمله S_i و جمله S_j است. طول هر الگوی تکراری پیداشده (pattern-length) برحسب ثانیه و طول جمله S_i و S_j برحسب ثانیه در این فرمول در نظر گرفته شده است. دلیل استفاده از طول الگو و طول جملات، نرمال‌کردن امتیازهای به‌دست آورده‌شده است. براساس این فرمول هرچه تعداد الگوهای تکراری پیداشده بین دو جمله بیشتر باشد، این دو جمله بیشتر شبیه هم هستند. در هر مرحله از تکرار الگوریتم، جمله‌ای که بیشترین امتیاز MMR را با توجه به فرمول (۱) گرفته باشد، به‌صورت حریصانه انتخاب شده و در خلاصه قرار می‌گیرد.

۲-۱-۳- روش TextRank

الگوریتم TextRank یک الگوریتم رتبه‌بندی مبتنی بر گراف و بی‌نظارت است. میزان اهمیت هر گره با توجه به اطلاعات سراسری که از کل گراف به‌صورت بازگشتی به‌دست می‌آید، محاسبه می‌شود. اگر G یک گراف جهت‌دار به‌صورت G=(V, E) باشد که V مجموعه‌ای از گره‌ها و E مجموعه‌ای از یال‌هاست. برای یک گره Vi، In(Vi) نشان‌دهندهٔ یال‌های ورودی به گره Vi است و Out(Vi) نشان‌دهندهٔ یال‌های خروجی از گره Vi است. برای خلاصه‌سازی، هر گره را برابر جمله و وزن روی یال‌ها میزان شباهت دو جمله با توجه به فرمول (۲) قرار داده شد؛ سپس امتیاز وزن‌دار (WS) هر گره با توجه به فرمول (۳) محاسبه شد [40].

$$WS(V_i) = (1-d) + d * \sum_{j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} WS(V_j) \quad (۳)$$

عدد d یک مقدار ثابت تعدیل[1] است که می‌تواند بین صفر و یک باشد. مقدار d به‌طورمعمول برابر ۰/۸۵ قرار داده می‌شود. در این فرمول، وزن بین گره V_i و V_j با w_{ji} نشان داده شده است که در اینجا میزان شباهت بین دو جمله را نشان می‌دهد. گراف ساخته‌شده یک گراف بی‌جهت است؛ بنابراین Out یک گره با In همان گره مساوی است.

الگوریتم از یک امتیاز دلخواه برای هر گره شروع و امتیاز هر گره براساس فرمول (۳) تا زمانی که الگوریتم همگرا شود، محاسبه می‌شود. معیار همگرایی نرخ خطا برای هر گره است که با توجه به فرمول (۴) محاسبه می‌شود [40].

$$S^{k+1}(V_i) - S^k(V_i) < T \quad (۴)$$

در این فرمول، k نشان‌دهندهٔ دفعات اجرا است؛ بدین معنی که اگر اختلاف امتیاز هر گره در دو تکرار متوالی

[1] Damping factor

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]
[DOI: 10.29252/jsdp.14.4.143]

از الگوریتم کمتر از مقدار آستانه‌ای T شود، الگوریتم به همگرایی رسیده و متوقف می‌شود. در آزمایش‌ها T برابر ۰/۰۰۱ قرار داده شد. امتیاز نهایی به‌دست‌آمده برای هر گره به مقدار اولیه تعیین‌شده برای هر گره وابستگی ندارد [40].

۲-۳- خلاصه‌سازی بانظارت

در یادگیری بانظارت، به‌ازای هر ورودی در داده‌های آموزشی، مقدار خروجی نیز مشخص است. هدف در یادگیری بانظارت، پیداکردن رابطه و تابع بین ورودی و خروجی است. در خلاصه‌سازی بانظارت، ابتدا یک مجموعهٔ آموزشی که شامل تعدادی جمله است، تهیه می‌شود. از هر جمله در مجموعه‌ی آموزشی یک بردار ویژگی استخراج و به هر بردار ویژگی برچسب بودن یا نبودن در خلاصه زده می‌شود. در مرحلهٔ بعد، از دو دسته‌بندی‌کنندهٔ ماشین بردار پشتیبان و بیز ساده استفاده شده است تا رابطهٔ بین ورودی‌ها و خروجی‌ها پیدا شود. از این رابطه برای برچسب‌زدن (برچسب خلاصه و غیرخلاصه) داده‌های جدید که در مجموعه آزمون وجود دارد، استفاده می‌شود.

۱-۲-۳- روش دسته‌بندی‌کنندهٔ ماشین بردار پشتیبان (SVM)

روش SVM جزو شاخهٔ روش‌های هسته‌ای[1] است و در حال حاضر جزو بهترین دسته‌بندی‌کننده‌های بانظارت به‌حساب می‌آید. ابتدا تعدادی ویژگی از هر جمله استخراج می‌شود. در پژوهش‌های مختلف ویژگی‌های متنوعی به‌منظور استخراج بهترین خلاصه از متن یا گفتار ارائه شده است. در اینجا سعی شد از ویژگی‌هایی که در مقالات مختلف نشان داده شده است و در خلاصه‌سازی مؤثر هستند، استفاده شود [41]، [8]، [7]، [33]. به‌عنوان مثال در خلاصه‌سازی اخبار نشان داده شده است که جملات نخست و جملات آخر هر سند دربرگیرندهٔ اطلاعات مفید و مهمی هستند که به‌طورمعمول توسط افراد خلاصه‌کننده در خلاصه قرار می‌گیرند. همچنین تعداد الگوهای کلیدی‌ای که در هر جمله وجود دارد، می‌تواند میزان اهمیت آن جمله را نشان دهد. این الگوهای کلیدی در گفتار با استفاده از الگوریتم پیشنهادشده در [42] استخراج شدند. امتیاز TextRank و امتیاز MMR که از مرحلهٔ قبل، خلاصه‌سازی بی‌نظارت، به‌دست آمده است نیز در این ویژگی‌ها گنجانده شد، تا تأثیر آنها مورد بررسی قرار بگیرد. ویژگی‌های استخراج‌شده از هر جمله در جدول (۱) نشان داده شده است. بنابراین، بعد از استخراج ویژگی از هر جمله، جمله با یک بردار ویژگی چهارده بعدی نشان داده می‌شود. بعد از استخراج ویژگی از هر جمله، جملات با استفاده از دسته‌بندی‌کنندهٔ SVM به دو دسته خلاصه و غیرخلاصه تقسیم می‌شوند. جملات با توجه به فاصلهٔشان از خط جداکننده که در دسته‌بندی‌کنندهٔ SVM تعیین شده است، امتیازدهی شده و به‌ترتیب در خلاصه قرار می‌گیرند. بدین ترتیب، ابتدا جملاتی که در دستهٔ خلاصه و در دورترین فاصله از خط جداکننده قرار گرفته‌اند در خلاصه قرار می‌گیرند؛ سپس، با توجه به نرخ خلاصه‌سازی اگر لازم بود جملاتی که در دسته غیرخلاصه قرار گرفته‌اند، اما در نزدیک‌ترین فاصله نسبت به خط جداکننده قرار دارند، در خلاصه قرار می‌گیرند.

(جدول-۱): ویژگی‌های استخراج‌شده برای خلاصه‌سازی
(Table-1): features used in our methods

	ویژگی‌ها
۱	طول جمله
۲	مکان جمله
۳	میزان شباهت جمله با سایر جملات (ماتریس شباهت)
۴	امتیاز TextRank
۵	امتیاز MMR
۶	جمله اول بودن
۷	میزان شباهت با جمله اول
۸	جمله دوم بودن
۹	میزان شباهت با جمله دوم
۱۰	جمله سوم بودن
۱۱	میزان شباهت با جمله سوم
۱۲	جمله آخر بودن
۱۳	میزان شباهت با جمله آخر
۱۴	تعداد الگوهای کلیدی در جمله

۲-۲-۳- روش دسته‌بندی‌کنندهٔ بیز ساده[2] (NB)

دسته‌بندی‌کنندهٔ بیز ساده جزو دسته‌بندی‌کننده‌های احتمالاتی است که براساس تئوری بیز با فرض مستقل‌بودن ویژگی‌ها عمل می‌کند. در این روش نیز از ویژگی‌های جدول (۱) و داده‌های برچسب‌داری که استفاده شده، در SVM

[1] Kernel Method

[2] Naïve Bayes

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]
[DOI: 10.29252/jsdp.14.4.143]

استفاده شد. داده‌های برچسب‌دار شامل بردار ویژگی استخراج‌شده از هر جمله و برچسب‌بودن و یا نبودن در خلاصه است. دسته‌بندی جملات به جملات خلاصه و غیرخلاصه با استفاده از الگوریتم NB انجام شده است. جملات با توجه به احتمال تعلق‌داشتن به دستهٔ خلاصه که توسط دسته‌بندی‌کنندهٔ NB تعیین شده است، رتبه‌بندی شده و به‌ترتیب در خلاصه قرار می‌گیرند.

۳-۳- خلاصه‌سازی نیمه‌نظارتی

در یادگیری نیمه‌نظارتی، مجموعه داده‌های آموزشی شامل تعداد کمی دادهٔ برچسب‌دار و تعداد بسیار زیادی داده بدون برچسب است. بنابراین در خلاصه‌سازی نیمه‌نظارتی، برای تعداد کمی از جملات، برچسب‌بودن و نبودن خلاصه مشخص است؛ اما برای تعداد بسیاری از جملات، این برچسب مشخص نیست. هدف این است که از داده‌های بدون برچسب نیز در آموزش دسته‌بندی‌کننده‌ها به امید پیداکردن دسته‌بندی بهتر استفاده شود. برای خلاصه‌سازی نیمه‌نظارتی از دو الگوریتم خودآموز و هم‌آموز استفاده شده است.

۱-۳-۳- الگوریتم نیمه‌نظارتی خودآموز با دسته‌بندی‌کنندهٔ SVM

در اینجا از الگوریتم نیمه‌نظارتی خودآموز با دسته‌بندی‌کنندهٔ SVM استفاده شده است. ابتدا SVM برروی داده‌های برچسب‌دار آموزش داده می‌شود؛ سپس داده‌های بدون برچسب با استفاده از این الگوریتم دسته‌بندی می‌شوند. داده‌هایی که با اطمینان بالایی دسته‌بندی شده باشند، به داده‌های آموزشی اضافه و سپس دوباره دسته‌بندی‌کننده SVM آموزش داده می‌شود. این کار تا جایی ادامه پیدا می‌کند که تمام داده‌های بدون برچسب، برچسب زده شوند. در SVM برای تعیین اینکه چه داده‌هایی با اطمینان بالایی دسته‌بندی شده‌اند، از فاصلهٔ داده‌ها تا خط جداکنندهٔ دو دسته استفاده می‌شود.

۲-۳-۳- الگوریتم نیمه‌نظارتی خودآموز با دسته‌بندی‌کنندهٔ NB

تنها تفاوت این روش با روش قبل این است که در آن از دسته‌بندی‌کنندهٔ NB استفاده شده است. برای تعیین اینکه چه داده‌هایی با اطمینان بالایی دسته‌بندی شده‌اند، از احتمال تعلق هر داده به دسته‌ها که با توجه به الگوریتم بیز ساده مشخص شده است، استفاده می‌شود.

مزیت الگوریتم خودآموز سادگی و عدم وابستگی به مدل دسته‌بندی‌کننده است؛ اما، ممکن است در مراحل یادگیری به‌اشتباه تقویت شود؛ بنابراین نسبت به داده‌های نوفه‌ای[1] حساس است؛ بنابراین، در ادامه الگوریتم هم‌آموز نیز مورد بررسی قرار گرفت.

۳-۳-۳- الگوریتم نیمه‌نظارتی هم‌آموز

در الگوریتم هم‌آموز، از دو دسته ویژگی مستقل از هم [43] و یا از دو دسته‌بندی‌کننده استفاده می‌شود [44]. بنابراین داده‌هایی که توسط یک دسته ویژگی یا یک دسته‌بندی‌کننده با احتمال پایینی دسته‌بندی می‌شوند، این شانس را دارند که توسط دستهٔ دیگر ویژگی‌ها یا دسته‌بندی‌کنندهٔ دوم به‌درستی دسته‌بندی شوند؛ چون تعداد ویژگی‌های استفاده‌شده برای آموزش (جدول) کم است و همچنین آنها را به دو دستهٔ مستقل نمی‌توان تقسیم کرد. در این مقاله از الگوریتم هم‌آموز با دو دسته‌بندی‌کننده استفاده شده است.

در این روش از ترکیبی از دسته‌بندی‌کنندهٔ SVM و NB استفاده شده است. ابتدا دسته‌بندی‌کنندهٔ NB با استفاده از داده‌های برچسب‌دار، آموزش داده می‌شود؛ سپس مدل تولیدشده برروی داده‌های آزمون اجرا می‌شود. داده‌هایی که با اطمینان بالایی دسته‌بندی شده‌اند به مجموعه آموزشی اضافه می‌شوند؛ سپس، دسته‌بندی‌کنندهٔ SVM با استفاده از مجموعه آموزشی جدید، آموزش داده می‌شود. داده‌هایی که با اطمینان بالایی دسته‌بندی شده‌اند، به مجموعه آموزشی اضافه می‌شوند. این روند ادامه پیدا می‌کند تا تمام داده‌های بدون برچسب دسته‌بندی شوند. مراحل اجرای الگوریتم در شکل (۱) نشان داده شده است. داده‌ی برچسب‌دار، جمله (بردار ویژگی استخراج‌شده از جمله) و برچسب بودن و یا نبودن آن جمله در خلاصه است. دادهٔ بدون برچسب تنها شامل جمله (بردار ویژگی استخراج‌شده از جمله) است.

۴- آزمایش‌ها و نتایج

مجموعهٔ خلاصه‌سازی استانداردی برای زبان فارسی در حوزهٔ گفتار وجود ندارد. بنابراین، یک مجموعه‌ای از اخبار خوانده‌شده توسط شانزده گوینده جمع‌آوری شد. برای تهیه این مجموعه از پیکرهٔ «پاسخ» استفاده شد. پیکرهٔ خلاصه‌ساز

[1] Outlier

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]
[DOI: 10.29252/jsdp.14.4.143]

فارسی «پاسخ» [45]، پیکره خلاصه‌سازی متون فارسی تهیه‌شده توسط دانشگاه فردوسی مشهد است. این پیکره براساس استاندارد DUC[1] [46] ساخته شده است. متون این مجموعه از سایت‌های خبری تابناک، پرس‌تی‌وی، فارس، ایرنا، همشهری، الف و جام‌جم جمع‌آوری شده است. موضوعات این متن‌ها شامل موضوعات اقتصادی، فرهنگی، اجتماعی، سیاسی، ورزشی و عملی می‌شود [45].

L مجموعه‌ی داده‌های برچسب‌دار است.
U مجموعه‌ی داده‌های بدون برچسب است.

تکرار: تا زمانی که همهٔ داده‌های بدون برچسب تمام شوند:
۱. دسته‌بندی‌کنندهٔ اول C_1 (NB) را برروی مجموعهٔ L آموزش بده.
۲. داده‌های U را با C_1 دسته‌بندی کن
۳. داده‌هایی را که با اطمینان بالایی دسته‌بندی شده‌اند، از مجموعه U حذف و به مجموعهٔ L اضافه کن.

سپس
۱. دسته‌بندی‌کنندهٔ دوم C_2 (SVM) را بر روی مجموعهٔ L آموزش بده.
۲. داده‌های U را با C_2 دسته‌بندی کن
۳. داده‌هایی را که با اطمینان بالایی دسته‌بندی شده‌اند، از مجموعه U حذف و به مجموعهٔ L اضافه کن.

(شکل-۱): الگوریتم هم‌آموز
(Figure-1): co-training algorithm

هشتاد متن مختلف از پیکرهٔ «پاسخ» به‌صورت تصادفی انتخاب و توسط شانزده گوینده، این متون ضبط و جمع‌آوری شد. از گویندگان آموزش‌دیده استفاده نشد؛ بنابراین برخی ناروانی‌ها مانند شروع مجدد و اشتباه در خواندن در این مجموعه وجود دارد. همچنین فایل‌های گفتاری شامل نوفهٔ محیط و نوفهٔ میکروفن است. وجود چنین خطاهایی باعث کاهش دقت خلاصه‌سازی نسبت به حالت تمیز می‌شود. به‌ازای هر گوینده، پنج متن مختلف ضبط شد. از بین پنج فایل ضبط‌شده به‌ازای هر گوینده، دو فایل برای آموزش و سه فایل برای آزمون خلاصه‌سازی قرار داده شد. در پیکرهٔ «پاسخ» هر متن شامل پنج خلاصه استخراجی تولیدشده توسط انسان است که به آنها خلاصه‌های مرجع گفته می‌شود. برای برچسب‌زدن به داده‌ها برای خلاصه‌سازی، به جملاتی که توسط بیشتر از سه نفر در خلاصه‌های مرجع قرار گرفته بودند، برچسب خلاصه و به سایر جملات برچسب غیرخلاصه زده شد. از این داده‌های برچسب‌دار برای آموزش دسته‌بندی‌کننده‌ها استفاده و برای ارزیابی خلاصه‌سازها از معیار ROUGE استفاده شد. این معیار مبتنی بر معیار فراخوان است. ابتدا در خلاصه‌سازی متن از این معیار استفاده می‌شد و بعدها در خلاصه‌سازی گفتار نیز مورد استفاده قرار گرفت [47]. معیار ROUGE، تعداد واحدهای هم‌پوشان بین خلاصهٔ خودکار تولیدشده و خلاصه‌های مرجع را شمارش می‌کند. ROUGE-n تعداد دنباله‌های nتایی[2] بین خلاصهٔ خودکار تولیدشده و خلاصه‌های مرجع را شمارش می‌کند. خروجی سامانه پیشنهادی بخش‌هایی از سیگنال گفتار هستند که در خلاصه قرار گرفته‌اند. برای اینکه بتوانیم خروجی سامانه پیشنهادی را با خلاصه‌های مرجع پیکره «پاسخ» با معیار ROUGE مقایسه کنیم، متن قسمت‌های گفتاری برای مقایسه با خلاصه‌های مرجع در نظر گرفته شد.

برای مقایسه روش‌های پیشنهادی از سامانه ایجاز [48] و دو روش پایه انتخاب نخستین جملات و طولانی‌ترین جملات نیز استفاده شد. ایجاز، یک سامانهٔ عملیاتی برای خلاصه‌سازی تک‌سندی متون خبری فارسی است که توسط دانشگاه فردوسی مشهد توسعه داده شده است [48]. در این سامانه بعد از نرمال‌سازی متن ورودی، جداسازی جملات، جداسازی کلمات، حذف هرز-واژه‌ها[3] و ریشه‌یابی کلمات، فاکتورهای مختلفی برای تعیین اهمیت جملات اعمال شده و پس از مرتب‌سازی جملات براساس وزن، جملات براساس نرخ فشرده‌سازی در خلاصه قرار می‌گیرند. از جمله فاکتورهایی که برای تعیین میزان اهمیت جملات در نظر گرفته شده شامل میزان شباهت با زمینه[4]، طول جمله، موقعیت جمله در متن، میزان شباهت با عنوان متن، تأثیر ضمایر، عبارات پراهمیت و یا خاص و غیره است. به‌طورمعمول نخستین جملات در متون خبری شامل شاخص‌ترین و مهم‌ترین بخش‌های خبری می‌باشند. همچنین طولانی‌ترین جملات شامل اطلاعات زیادی هستند. از این رو از این دو روش نیز به‌عنوان روش‌های پایه برای ارزیابی خلاصه‌ساز استفاده شد. ابتدا، فایل‌های گفتاری به‌صورت دستی جمله‌بندی شدند تا خطاهای تعیین انتهای جملات در خلاصه‌سازی تأثیری نگذارد. در کارهای آینده باید تأثیر جمله‌بندی خودکار با استفاده از روش‌های پیشنهادی همچون [49] مورد بررسی قرار بگیرد. از بین

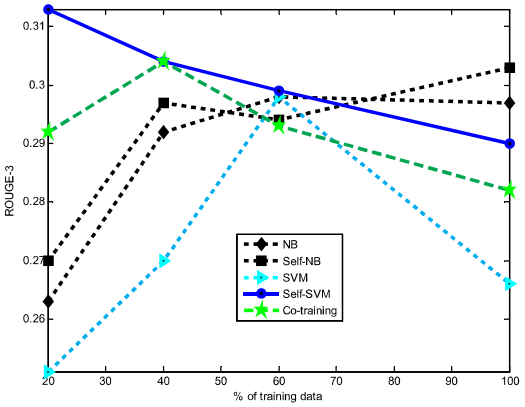
[1] Document Understanding Conference

[2] N-gram
[3] Stop words
[4] Context

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]
[DOI: 10.29252/jsdp.14.4.143]

جملات برچسب‌زده‌شده، ۴۵۲ جمله در مجموعه آموزشی و ۵۶۰ جمله در مجموعه آزمون قرار داده شد.

در آزمایش نخست، به‌ازای تعداد داده‌های آموزشی مختلف مدل‌ها را آموزش دادیم تا تأثیر تعداد داده‌های آموزشی را بررسی کنیم. در ابتدا به‌ازای تمام ویژگی‌ها، هر پنج الگوریتم ارائه‌شده در این مقاله در شکل (۲) با یکدیگر مقایسه شدند.



(شکل-۲): بررسی تاثیر تعداد داده‌های آموزشی مختلف برروی روش‌های مختلف به ازای تمام ویژگی‌ها برحسب معیار ROUGE-3

(Figure-2): comparison of different methods based on ROUGE-3 for different amount of training data using all features

همان‌طور که در شکل (۲) مشخص است، زمانی که تنها از ۴۰٪ داده‌های آموزشی استفاده شده است، به‌ترتیب روش خودآموز SVM، روش هم‌آموز و روش خودآموز NB بهترین دقت را داشته‌اند. حتی زمانی که از ۱۰۰٪ داده‌های آموزشی استفاده شده روش نیمه‌نظارتی خودآموز NB بهترین کارایی را داشته است؛ زیرا تعداد کل داده‌های آموزشی کم بوده و الگوریتم نیمه‌نظارتی توانسته است کارایی بهتری را حتی نسبت به روش‌های بانظارت به‌دست آورد. نـــکته‌ای که در این شکل وجود دارد، این است که با افزایش داده‌های آموزشی، دقت روش‌ها به‌طور صعودی افزایش پیدا نکرده است؛ درحالی‌که انتظار می‌رود با افزایش داده‌ها دقت دسته‌بندی‌کننده‌ها روندی صعودی داشته باشد. به نظر می‌رسد که دلیل این امر وجود ویژگی‌های نوفه‌ای است. دسته‌بندی‌کننده‌ها نتوانسته‌اند با وجود هر چهارده ویژگی جدول (۱) داده‌ها را به‌خوبی دسته‌بندی کنند. بنابراین، در مرحلهٔ بعد آزمایشی انجام شد تا تأثیر هر یک ویژگی‌ها را مورد بررسی قرار دهد.

برای خلاصه‌سازی با نرخ ۳۰٪ و با استفاده از الگوریتم هم‌آموز و با استفاده از الگوریتم انتخاب ویژگی رو به جلو ویژگی‌های انتخاب‌شده در هر مرحله در جدول (۲) نشان داده شده است.

(جدول-۲): بررسی تاثیر ویژگی‌ها برروی الگوریتم هم‌آموز

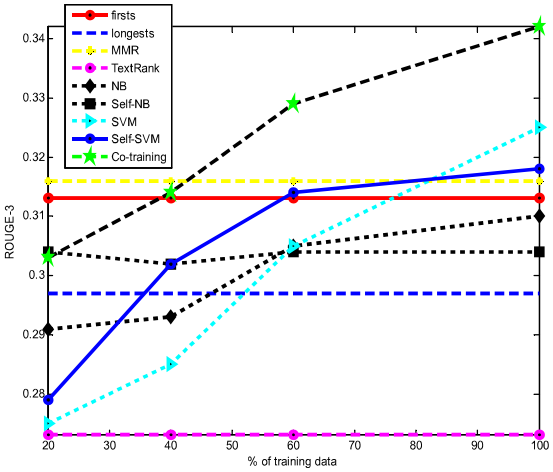
(Table-2): feature comparison for co-training algorithm

ویژگی	ROUGE-3 (%)
امتیاز MMR	۳۱/۶
+ مکان جمله	۳۳/۱
+ میزان شباهت با جمله اول	۳۳/۴
+ میزان شباهت با جمله دوم	**۳۴/۲**
+ میزان شباهت با جمله سوم	۳۳/۹
+ امتیاز TextRank	۳۳/۵
+ جمله اول بودن	۳۲/۶
+ جمله آخر بودن	۳۲/۳
+ میزان شباهت جمله با سایر جملات (ماتریس شباهت)	۳۱/۴
+ تعداد الگوهایی کلیدی در جمله	۳۱/۶
+ جمله دوم بودن	۳۰
+ طول جمله	۳۰/۲
+ میزان شباهت با جمله آخر	۲۸/۲

همان‌طورکه در جدول (۲) مشخص است با درنظرگرفتن چهار ویژگی ۵، ۲، ۷، ۹ از جدول (۱) یعنی ویژگی‌های امتیاز MMR، مکان جمله، میزان شباهت با جمله نخست و میزان شباهت با جمله دوم، الگوریتم هم‌آموز با نرخ خلاصه‌سازی ۳۰٪ با معیار ROUGE-3 بهترین کارایی رسیده است. البته می‌دانیم که الگوریتم انتخاب ویژگی رو به جلو، یک الگوریتم حریصانه است؛ بنابراین ممکن است، بهترین زیرمجموعه از ویژگی‌ها انتخاب نشوند. همچنین ممکن است، به‌ازای الگوریتم‌های مختلف و با درنظرگرفتن نرخ‌های مختلف خلاصه‌سازی ترتیب بهترین ویژگی‌ها تغییر کند. این موارد در کارهای آینده باید مورد بررسی قرار گیرد؛ اما با درنظرگرفتن تنها ویژگی‌های ۵، ۲، ۷، ۹ از جدول (۱) و با معیار ROUGE-3 کارایی هر هفت الگوریتم ارائه‌شده در این مقاله به‌همراه دو روش پایهٔ انتخاب نخستین جملات و انتخاب طولانی‌ترین جملات در شکل (۳) نشان داده شده است.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.143]



(شکل-۳): مقایسه کارایی روش‌های مختلف برحسب معیار ROUGE-3 به‌ازای درصد استفاده‌شده از داده‌های آموزشی

(Figure-3): performance comparison between different methods based on different amount of training data

همان‌طورکه در شکل (۳) مشخص است، در این حالت با افزایش داده‌های آموزشی کارایی روش‌های بانظارت و نیمه‌نظارتی روند افزایشی پیدا کرده و در همهٔ موارد کارایی الگوریتم نیمه‌نظارتی هم‌آموز از روش‌های بانظارت و دیگر روش‌های نیمه‌نظارتی بهتر بوده است. این نتایج نشان می‌دهد، زمانی که داده‌های برچسب‌دار کمی در اختیار داریم، استفاده از روش‌های نیمه‌نظارتی و کمک‌گرفتن از داده‌های بدون برچسب کارایی سامانه را می‌تواند افزایش دهد.

نتایج دو روش پایه، انتخاب نخستین جملات و انتخاب طولانی‌ترین جملات نیز در شکل (۳) نشان داده شده است. چون این دو روش به تعداد داده‌های آموزشی وابسته نیستند، به‌صورت یک خط افقی در شکل نمایان شده‌اند. الگوریتم هم‌آموز توانسته است به‌ازای داده‌های آموزشی مختلف همواره بهتر از طولانی‌ترین جمله عمل کند و زمانی که بیشتر از ۴۰٪ از داده‌های آموزشی استفاده شده است از انتخاب نخستین جملات نیز عملکرد بهتری داشته است. زمانی که تعداد داده‌های آموزشی بسیار کم بوده است (۲۰٪ تعداد کل داده‌های آموزشی) روش بی‌نظارت MMR نسبت به تمام روش‌ها کارایی بهتری داشته است. در مجموعهٔ آموزشی، تعداد داده‌هایی که برچسب خلاصه دارند از تعداد داده‌های برچسب غیرخلاصه دارند، خیلی کمتر است. این نامتوازن‌بودن دسته‌ها در کارایی دسته‌بندی‌کننده‌ها می‌تواند تأثیرگذار باشد. در زمانی که تعداد داده‌های آموزشی بسیار کم است، ممکن است نامتوازن بودن دسته‌ها تأثیر منفی بیشتری برروی دسته‌بندی‌کننده‌ها گذاشته باشد. درنتیجه، کارایی روش‌های بانظارت و نیمه‌نظارتی از روش بی‌نظارت کمتر شده است. بنابراین، بهتر است در کارهای آینده از روش‌های نمونه‌برداری که برای متوازن‌کردن دسته‌ها به کار می‌رود، استفاده و تأثیر متوازن‌کردن دسته‌ها برروی روش‌های بانظارت و نیمه‌نظارتی بررسی شود.

در همهٔ حالات، الگوریتم بی‌نظارت TextRank کارایی کمتری نسبت به الگوریتم بی‌نظارت MMR داشته است. دلیل این امر هم می‌تواند این باشد که در الگوریتم MMR، سازوکاری برای حذف افزونگی در نظر گرفته شده است. درحالی‌که در روش TextRank این سازوکار وجود نداشته و جملات تنها براساس شباهتشان به یکدیگر رتبه‌بندی می‌شوند. در روش TextRank نیز با تعریف یک مقدار آستانه از قرارگرفتن جملاتی که شبیه به یکدیگر هستند در خلاصه می‌توان جلوگیری کرد. مقایسهٔ دو الگوریتم بانظارت SVM و نیمه‌نظارتی SVM خودآموز نشان می‌دهد، در زمانی که تعداد داده‌های آموزشی کمی در اختیار بوده است، طبق انتظار روش نیمه‌نظارتی عملکرد بهتری نسبت به روش بانظارت داشته است. چنین نکته‌ای در مورد در الگوریتم NB و NB خودآموز نیز صادق است.

در آزمایش سوم، به‌ازای نرخ‌های مختلف خلاصه‌سازی روش‌های مختلف با یکدیگر مقایسه شده‌اند. روشی که در هر نرخ خلاصه‌سازی بیشترین کارایی را داشته با علامت ستاره در شکل (۴) نشان داده شده است. آزمایش‌ها با چهار ویژگی ۵،۲، ۷، ۹ انجام شده است. به‌ازای نرخ ۷۰٪، الگوریتم خودآموز NB، نرخ ۵۰٪ الگوریتم خودآموز SVM، نرخ ۳۰٪ الگوریتم هم‌آموز، نرخ ۱۰٪ الگوریتم MMR با معیار ROUGE-3 بهترین نتایج را کسب کرده‌اند. بنابراین به‌ازای نرخ‌های مختلف خلاصه‌سازی ممکن است روش‌هایی متفاوتی کارا باشند. البته در همهٔ موارد روش‌های پیشنهادی کارایی بهتری نسبت به دو روش پایهٔ انتخاب نخستین جملات و انتخاب طولانی‌ترین جملات داشته‌اند.

در آزمایش آخر، کارایی روش‌های مختلف به‌ازای نرخ خلاصه‌سازی ۳۰٪ با چهار ویژگی ۵،۲، ۷، ۹ با معیارهای ROUGE-1، ROUGE-2 و ROUGE-3 در جدول (۳) نشان داده شده است.

الگوریتم‌های ارائه‌شده در این مقاله با دو روش پایه

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

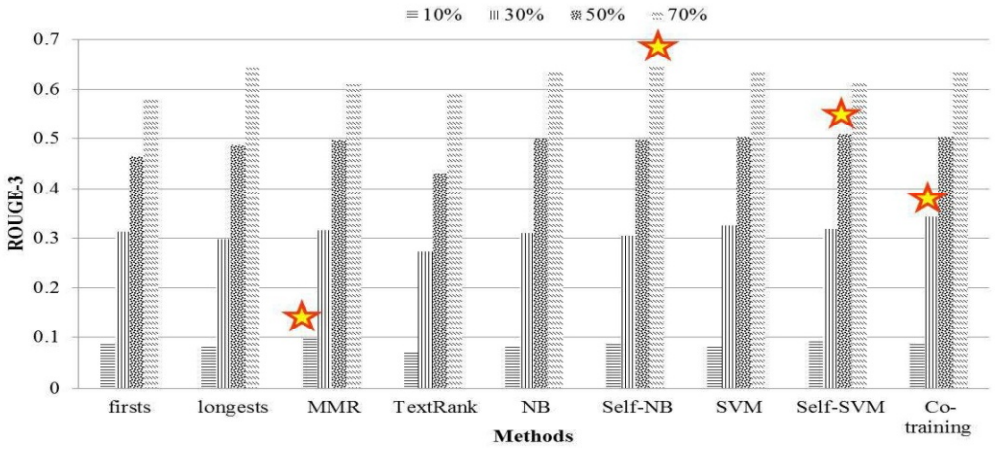
[DOI: 10.29252/jsdp.14.4.143]

انتخاب نخستین جملات و انتخاب طولانی‌ترین جملات و همچنین سامانهٔ ایجاز مقایسه شده‌اند و نتایج در جدول (۳) نشان داده شده است. الگوریتم هم‌آموز توانسته بهترین کارایی را در بین تمام روش‌ها داشته باشد. همان‌طورکه درقبل نیز اشاره شد، الگوریتم TextRank کمترین کارایی را در بین روش‌های پیشنهادی داشته است. روش هم‌آموز توانسته است، بدون استفاده از رونوشت، تنها با استفاده از سیگنال گفتار بهتر از سامانهٔ خلاصه‌سازی برروی متن (ایجاز) عمل کند.

(جدول-۳): مقایسهٔ کارایی روش‌های مختلف خلاصه‌سازی به‌ازای نرخ خلاصه‌سازی ۳۰٪ با چهار ویژگی ۲، ۵، ۷، ۹

(Table-3): performance comparison of different methods using features 2, 5, 7, 9 with summarization rate of 30%

روش	ROUGE-1 (%)	ROUGE-2 (%)	ROUGE-3 (%)
اولین جملات	۵۸/۲	۳۶	۳۱/۳
طولانی‌ترین جملات	۵۷/۳	۳۴/۳	۲۹/۷
ایجاز	۵۰/۱	۲۸	۲۱/۹
MMR	۵۹/۷	۳۶/۶	۳۱/۶
TextRank	۵۵/۸	۳۲	۲۷/۳
NB	۵۸/۷	۳۵/۹	۳۱
Self-NB	۵۸/۲	۳۵/۳	۳۰/۴
SVM	۵۹/۷	۳۷/۴	۳۲/۵
Self-SVM	۵۹/۳	۳۶/۶	۳۱/۸
Co-training	۶۱/۲	۳۹/۲	۳۴/۲



(شکل-۴): مقایسهٔ کارایی روش‌ها به‌ازای نرخ‌های مختلف خلاصه‌سازی

(Figure-4): performance comparison of different methods using various summarization rates

۵- نتیجه‌گیری و کارهای آینده

در این مقاله خلاصه‌سازی گفتار بدون استفاده از رونوشت مورد بررسی قرار گرفت. برای استخراج الگوهای تکراری بین هر دو جملهٔ گفتاری از الگوریتم S-DTW استفاده شد؛ سپس روش‌های مختلف یادگیری ماشین، بانظارت، بی‌نظارت و نیمه‌نظارتی مورد آزمایش قرار داده شد. خلاصه‌سازی با نرخ‌های مختلف خلاصه‌سازی و تعداد داده‌های مختلف و با ویژگی‌های متفاوت بررسی شد. نتایج نشان داده شد که بدون استفاده از رونوشت به خلاصه‌ای بهتر از حالت‌های پایه (انتخاب نخستین جملات، انتخاب طولانی‌ترین جملات) می‌توان رسید. در آزمایش‌ها نشان داده شد، انتخاب ویژگی‌های مناسب تأثیر قابل توجه در نتایج نهایی می‌گذارد. همچنین در زمانی که تعداد داده‌های آموزشی کمی در اختیار است، استفاده از روش‌های نیمه‌نظارتی می‌تواند باعث بهبود نتایج شود.

در ادامه، این روش خلاصه‌سازی باید با تعیین انتهای جملات به‌صورت خودکار آزمایش شوند. آزمایش‌ها برروی پیکره‌های بزرگ‌تر زبان فارسی و پیکره به زبان انگلیسی نیز انجام شود. پیچیدگی زمانی الگوریتم مورد بررسی قرار بگیرد

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]
[DOI: 10.29252/jsdp.14.4.143]

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.143]

و از الگوریتم‌های بهبودیافته S-DTW استفاده شود. کارایی الگوریتم با داشتن چندین گوینده در شرایط آکوستیکی سخت مورد آزمایش قرارگیرند.

6-References ۶- مراجع

[1] A. McCallum, "An Ecologically Valid Evaluation of Speech Summarizationin the University Lecture Domain," MSc thesis, University of Toronto, 2012.

[2] S. R. Maskey, "Automatic Broadcast News Speech Summarization," PhD thesis, School of Arts and Sciences, Columbia University, 2008.

[3] R. Flamary, X. Anguera, and N. Oliver, "Spoken WordCloud: Clustering recurrent patterns in speech," in *CBMI* 2011, pp. 133-138.

[4] Y. Liu, S. Xie, and F. Liu, "Using n-best recognition output for extractive summarization and keyword extraction in meeting speech," in *ICASSP* 2010, pp. 5310-5313.

[5] S. Xie and Y. Liu, "Using N-Best Lists and Confusion Networks for Meeting Summarization," *IEEE Transactions on Audio, Speech & Language Processing,* vol. 19, no. 5, pp. 1160-1169, 2011.

[6] J. Zhang, R. H. Y. Chan, P. Fung, and L. Cao, "A comparative study on speech summarization of broadcast news and lecture speech," in *INTERSPEECH* 2007, pp. 2781-2784.

[7] S. Xie, D. Hakkani-Tür, B. Favre, and Y. Liu, "Integrating prosodic features in extractive meeting summarization," in *ASRU*, Merano/Meran, Italy, 2009, pp. 387-391.

[8] S. Xie, Y. Liu, and H. Lin, "Evaluating the effectiveness of features and sampling in extractive meeting summarization," presented at the SLT 2008.

[9] S. Xie and Y. Liu, "Improving supervised learning for meeting summarization using sampling and regression," *Computer Speech & Language,* vol. 24, no. 3, pp. 495-514, 2010.

[10] S.-H. Lin and B. Chen, "A Risk Minimization Framework for Extractive Speech Summarization," in *ACL* Uppsala, Sweden, 2010, pp. 79-87.

[11] B. Chen and S.-H. Lin, "A Risk-Aware Modeling Framework for Speech Summarization," *IEEE Transactions on Audio, Speech & Language Processing,* vol. 20, no. 1, pp. 211-222, 2012.

[12] J. J. Zhang and P. Fung, "Learning deep rhetorical structure for extractive speech summarization," in *ICASSP*, 2010, pp. 5302-5305.

[13] J. Zhang, H. Yuan, and X. Pan, "rhetorical-state SVM for Lecture speech summarization," *Information Technology Journal,* 2014.

[14] S.-H. Lin, Y.-M. Yeh, and B. Chen, "Leveraging Kullback–Leibler Divergence Measures and Information-Rich Cues for Speech Summarization," *IEEE Transactions on Audio, Speech, and Language,* vol. 19 no. 4, pp. 871-882, May 2011.

[15] S. Xie, H. Lin, and Y. Liu, "Semi-supervised extractive speech summarization via co-training algorithm," in *INTERSPEECH* 2010, pp. 2522-2525.

[16] B. Chen, H.-C. Chang, and K.-Y. Chen, "Sentence modeling for extractive speech summarization," in *ICME*, San Jose, CA, USA, 2013, pp. 1-6.

[17] B. Chen, S.-H. Lin, Y.-M. Chang, and J.-W. Liu, "Extractive speech summarization using evaluation metric-related training criteria," *Information Processing and Management,* vol. 49, no. 1, pp. 1-12, 2013.

[18] D. Gillick, K. Riedhammer, B. Favre, and D. Z. Hakkani-Tür, "A global optimization framework for meeting summarization," in *ICASSP* 2009, pp. 4769-4772.

[19] K. Riedhammer, B. Favre, and D. Hakkani-Tür, "Long story short - Global unsupervised models for keyphrase based meeting summarization," *Speech Communication,* vol. 52, pp. 801-815, 2010.

[20] Y.-N. Chen, Y. Huang, C.-f. Yeh, and L.-S. Lee, "Spoken Lecture Summarization by Random Walk over a Graph Constructed with Automatically Extracted Key Terms," in *INTERSPEECH* 2011, pp. 933-936.

[21] T. J. Hazen, "Latent Topic Modeling for Audio Corpus Summarization," in *INTERSPEECH* 2011, pp. 913-916.

[22] L. Wang and C. Cardie, "Unsupervised Topic Modeling Approaches to Decision Summarization in Spoken Meetings," in *SIGDIAL Conference*, 2012, pp. 40-49.

[23] K.-Y. Chen *et al.*, "Extractive Broadcast News Summarization Leveraging Recurrent Neural Network Language Modeling Techniques," *IEEE/ACM Transactions on Audio, Speech & Language Process-ing,* vol. 23, no. 8, pp. 1322-1334, 2015.

[24] M. H. Bokaei, H. Sameti, and Y. Liu, "Extractive summarization of multi-party meetings through discourse segmentation," *Natural Language Engineering,* vol. 22, no. 1, pp. 41-72, 2016.

[25] M.-H. Siu, H. Gish, A. Chan, W. Belfield, and S. Lowe, "Unsupervised training of an HMM-based self-organizing unit recognizer with applications to topic classification and keyword discovery," *Computer Speech & Language,* vol. 28, no. 1, pp. 210-223, 2014.

[26] N. F. Chen, B. Ma, and H. Li, "Minimal-resource phonetic language models to summarize untranscribed speech," in *ICASSP* 2013, pp. 8357-8361.

[27] A. Muscariello, G. Gravier, and F. Bimbot, "Unsupervised Motif Acquisition in Speech via Seeded Discovery and Template Matching Combination," *IEEE Transactions on Audio, Speech & Language Processing,* vol. 20, no. 7, pp. 2031-2044, 2012.

[28] J. R. Glass, "Towards unsupervised speech processing," in *ISSPA* Montreal, QC, Canada, 2012, pp. 1-4.

[29] D. F. Harwath, T. J. Hazen, and J. R. Glass, "Zero resource spoken audio corpus analysis," in *ICASSP* 2013, pp. 8555-8559.

[30] S. Maskey and J. Hirschberg, "Summarizing Speech Without Text Using Hidden Markov Models," presented at the HLT-NAACL, 2006.

[31] S. H. Yella, V. Varma, and K. Prahallad, "Prominence based scoring of speech segments for automatic speech-to-speech summarization," in *INTERSPEECH* 2010, pp. 1297-1300.

[32] S. K. Jauhar, Y.-N. Chen, and F. Metze, "Prosody-Based Unsupervised Speech Summarization with Two-Layer Mutually Reinforced Random Walk," in *IJCNLP*, Nagoya, Japan, 2013, pp. 648-654.

[33] J. Zhang and H. Yuan, "Speech Summarization without Lexical Features for Mandarin Presentation Speech," in *IALP*, Urumqi, China, 2013, pp. 147-150.

[34] X. Zhu, G. Penn, and F. Rudzicz, "Summarizing multiple spoken documents: finding evidence - from untranscribed audio," in *ACL/IJCNLP*, 2009, pp. 549-557.

[35] A. S. Park and J. R. Glass, "Unsupervised Pattern Discovery in Speech," *IEEE Transactions on Audio, Speech & Language Processing,* vol. 16, no. 1, pp. 186-197, 2008.

[36] A. Jansen, K. Church, and H. Hermansky, "Towards spoken term discovery at scale with zero resources," in *INTERSPEECH* 2010, pp. 1676-1679.

[37] Y. Zhang, "Unsupervised Speech Processing with Applications to Query-by-Example Spoken Term Detection," PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2013.

[38] Y. Zhang and J. R. Glass, "Towards multi-speaker unsupervised speech pattern discovery," in *ICASSP* 2010, pp. 4366-4369.

[39] D. F. Harwath, "Unsupervised Modeling of Latent Topics and Lexical Units in Speech Audio," MSc thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2013.

[40] R. Mihalcea and P. Tarau, "TextRank: Bringing Order into Text," in *EMNLP* Barcelona, Spain, 2004, pp. 404-411.

[41] S. Maskey and J. Hirschberg, "Comparing Lexial, Acoustic/Prosodic, Discourse and Structural Features for Speech Summarization," in *Eurospeech* Lisbon, Portugal, 2005.

[42] H. S. Jafari and M. M. Homayounpour, "key pattern recognition from Persian speech signal without transcript," presented at the 19th National CSI Computer Conference, Shahid Beheshti University, Tehran, Iran, 2014.

[43] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *11th Annual Conference on Computational Learning Theory*, 1998, pp. 92-100.

[44] S. A. Goldman and Y. Zhou, "Enhancing Supervised Learning with Unlabeled Data," in *ICML*, Stanford, CA, USA, 2000, pp. 327-334.

[45] B. B. Moghaddas, M. Kahani, S. A. Toosi, AsefPourmasoumi, and A. Estiri, "Pasokh: A standard corpus for the evaluation of Persian text summarizers," in *ICCKE*, Mashhad, Iran, 2013, pp. 471-475.

[46] DUC. (2013). *http://duc.nist.gov/.*

[47] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries. In Proceedings," in *workshop on text summarization branches out, 2004, pp. 25-26.*

[۴۸] آ. پورمعصومی، م. کاهانی، س. ا. طوسی، ا. استیری، ه. قائمی، (۱۳۹۳). ایجاز: یک سامانه عملیاتی برای خلاصه‌سازی تک سندی متون خبری فارسی. پردازش علائم و داده‌ها ۱ (۲۱): ۴۸-۳۳.

[48] A. pourmasoomi, M. kahani, S. A. Toosi, and A. Estiri, "Ijaz: An Operational system for single-document summarization of Persian news texts," *JSDP,* vol. 11, no. 1, pp. 33-48, 2014.

[۴۹] ه. س جعفری و م.م همایون‌پور، (۱۳۹۲). تشخیص الگوهای کلیدی از سیگنال گفتار فارسی بدون استفاده از رونوشت. نوزدهمین کنفرانس ملی سالانه انجمن کامپیوتر ایران. دانشگاه شهید بهشتی، تهران، ایران.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.143]

[49] H. S. Jafari and M. M. Homayounpour, "Persian speech sentence segmentation without speech recognition," presented at the Iranian Conference on Intelligent Systems (ICIS), Bam, Kerman, 2014.



هـدی‌سادات جعفری در سال ۱۳۸۹ دورهٔ کارشناسی خود را رشتهٔ مهندسی کامپیوتر در دانشگاه شهید بهشتی به پایان رسانده و در سال ۱۳۹۲ مدرک کارشناسی‌ارشد خود را در همان رشته از دانشگاه صنعتی امیرکبیر اخذ کرده است. زمینه‌های مورد علاقه ایشان بازشناسی گفتار، پردازش زبان طبیعی، داده‌کاوی و هستان‌شناسی است.

نشانه رایانامهٔ ایشان عبارت است از:

hsjafari@aut.ac.ir



محمد مهدی همایون‌پور تحصیلات خود را در مقطع کـــارشناسی در رشته مهندسی برق (الکترونیک) در دانشگاه صنعتی امیرکبیر (سال ۱۳۶۶) و کارشناسی ارشد را در رشته برق (مخابرات)، از دانشگاه خواجـه‌نصیرالدین‌طـوسی (سال ۱۳۹۶) اخـذ کرد. ایشان کارشناســـی ارشـد دوم خود را در زمـــینه فونتیک (۱۳۷۴) در دانشگاه سوربون جدید فـرانســـه و هم‌زمان دورهٔ دکترای خود را در دانشگاه پاریس ۱۱ در زمینه مهندسی برق (۱۳۷۴) به‌پایان رسانـد. وی از ســـال ۱۳۷۴ در سمت عضو هیأت علمی دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر به تدریس و پژوهش مشغول است. زمینه‌های تخصصی مورد علاقه ایشان شامل پردازش سیگنال‌های رقمی، پردازش گفتار، یادگیری عمیق، پردازش زبان طبیعی, تشخیص نفوذ در سامانه‌ها و شبکه‌های رایانه‌ای، اتوماسیون صنعتی، چندرسانه‌ای و طراحی سخت‌افزار است.

نشانه رایانامهٔ ایشان عبارت است از:

homayoun@aut.ac.ir

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.143]