

(Research Article)

Investigating the performance of machine learning-based methods in classroom reverberation time estimation using neural networks

M. Shafieian*, S. Nouri Zehab

Faculty of Media Engineering and Technology, IRIB University

Received: 2021/11/20, Accepted: 2022/09/04

DOR 20.1001.1.23455748.1401.10.1.6.5

Abstract

Classrooms, as one of the most important educational environments, play a major role in the learning and academic progress of students. reverberation time, as one of the most important acoustic parameters inside rooms, has a significant effect on sound quality. The inefficiency of classical formulas such as Sabin, caused this article to examine the use of machine learning methods as an alternative method for predicting the environment's reverberation time. In this research, firstly, by using methods based on geometrical acoustics and by using Odeon software, the collection of required data sets at frequencies of 500 and 2000 Hz is done. In this dataset, 4 classrooms with a rectangular space, along with elements such as desks and chairs, windows, and doors, were used. After that, to provide a system based on machine learning, multilayer perceptron neural network and neural network based on radial basis functions along with K-means clustering algorithm and also convolutional neural network has been used. These models consider the characteristics of the environment and finally estimate the values of reverberation time as a function of frequency. In this research, by using the multi-layer perceptron neural network, the determination coefficient was 93% for the frequency of 500 Hz and 95% for the frequency of 2000 Hz. Also, by using the neural network based on radial basis functions, for the frequency of 500 Hz, the coefficient of determination was 82% and for the frequency of 2000 Hz, the coefficient of determination was recorded as 89%. Also, by using a one-dimensional convolutional neural network, a determination coefficient of 94% was recorded for the frequency of 500 Hz, and a determination coefficient of 96% for the frequency of 2000 Hz.

Keywords: Classroom acoustic, Reverberation time, Multilayer perceptrons, Radial basis function, 1D Convolutional.

pp. 54-66 (In Persian)

* Corresponding author E-mail: noghani@eng.ikiu.ac.ir

بررسی عملکرد روش‌های مبتنی بر یادگیری ماشین در تخمین زمان واخنش کلاس درس با استفاده از شبکه‌های عصبی

معصومه شفیعیان^{*}، سلمان نوری زه‌آب

دانشکده فنی مهندسی رسانه، دانشگاه صداوسیما

دریافت: ۱۴۰۰/۰۸/۲۹، پذیرش: ۱۴۰۱/۰۶/۱۳

چکیده

کلاس‌های درس به‌عنوان یکی از مهم‌ترین محیط‌های آموزشی نقش عمده‌ای در یادگیری و پیشرفت تحصیلی دانش‌آموزان دارند. زمان واخنش به‌عنوان یکی از مهم‌ترین شبه‌سنج‌های صوتی در داخل اتاق‌ها، تأثیر بسزایی در کیفیت صدا دارد. عدم کارایی مناسب فرمول‌های کلاسیک مانند سابین، باعث شد که در این مقاله به بررسی استفاده از روش‌های یادگیری ماشین به‌عنوان یک روش جایگزین برای پیش‌بینی زمان واخنش محیط پرداخته شود. در این پژوهش ابتدا با استفاده از روش‌های مبتنی بر صوتیات هندسی و با استفاده از نرم‌افزار اودئون به جمع‌آوری مجموعه داده‌گان مورد نیاز در بسامدهای ۵۰۰ و ۲۰۰۰ هرتز پرداخته می‌شود. در این مجموعه داده‌گان چهار کلاس درس با فضایی مستطیل شکل، همراه با عنصرهایی مانند میز و صندلی و پنجره و در، استفاده شد. پس از آن به‌منظور ارائه یک سامانه مبتنی بر یادگیری ماشین از شبکه عصبی پرسپترون چندلایه و شبکه عصبی مبتنی بر توابع پایه شعاعی به همراه الگوریتم خوشه‌بندی کی-میانی و هم‌چنین شبکه عصبی کانولوشن استفاده شده است. این الگوها ویژگی‌های محیط را در نظر می‌گیرند و در نهایت مقادیر زمان واخنش را به‌عنوان تابعی از بسامد برآورد می‌کنند. در این پژوهش با استفاده از شبکه عصبی پرسپترون چندلایه برای بسامد ۵۰۰ هرتز، ضریب تعیین ۹۳ درصد و برای بسامد ۲۰۰۰ هرتز، ضریب تعیین ۹۵ درصد حاصل شد. هم‌چنین با استفاده از شبکه عصبی مبتنی بر توابع پایه شعاعی برای بسامد ۵۰۰ هرتز، ضریب تعیین ۸۲ درصد و برای بسامد ۲۰۰۰ هرتز، ضریب تعیین ۸۹ درصد ثبت شد. هم‌چنین با استفاده از شبکه عصبی کانولوشن یک-بعدی برای بسامد ۵۰۰ هرتز، ضریب تعیین ۹۴ درصد و برای بسامد ۲۰۰۰ هرتز ضریب تعیین ۹۶ درصد ثبت شد.

کلیدواژه‌ها: صوتیات کلاس درس، زمان واخنش، پرسپترون چندلایه، توابع پایه شعاعی، کانولوشن یک-بعدی.

۱. مقدمه

در محیط‌های آموزشی بخش عمده‌ای از فرآیند آموزش از طریق درست شنیدن اتفاق می‌افتد. تمرکز و یادگیری به محیط بهینه صوتی نیاز دارد، و شرایط صوتی نامناسب باعث اختلال در امر یادگیری دانش‌آموزان می‌شود [۱]. یکی از شبه‌سنج‌های مهم به منظور توصیف ویژگی‌های صوتی محیط، زمان واخنش است. سابین^۱ اولین نفری بود که به پیش‌بینی زمان واخنش پرداخت و با توجه به تحقیقات تجربی، به رابطه‌ای برای زمان واخنش رسید که نشان می‌داد زمان واخنش متأثر از حجم محیط و ضریب

جذب است که در معادله (۱) بیان می‌شود:

$$T_{\text{sabin}} = \frac{0.16 * V}{S * \bar{\alpha}_{\text{SAB}}} \quad (1)$$

$$\bar{\alpha}_{\text{SAB}} = \frac{1}{S} \sum_i S_i * \alpha_i \quad (2)$$

که در این رابطه، V حجم اتاق است و S مساحت کل سطوح می‌باشند و S_i و α_i به ترتیب مساحت و ضریب جذب سطح شماره i -ام محیط می‌باشند. همان‌طور که معلوم است این معادلات کلاسیک، فرمول‌های نسبتاً ساده‌ای هستند و مهم‌ترین فرض این معادلات این است که میدان صوتی در هر محیط محصور به‌صورت پخشا باشد و اگر میدان صوتی یک محیط پخشا نباشد زمان واخنش در یک محیط نقطه به نقطه متفاوت خواهد بود و

^{*} نویسنده پاسخگو: shafieian@iribu.ac.ir

^۱ Sabin

زمان واخنش می‌گویند و در تعریف به معنای مدت زمانی است که بعد از خاموش شدن منبع صدا، طول می‌کشد تا تراز فشار صوتی به میزان ۶۰ دسی‌بل کاهش پیدا کند. این شبه‌سنج را با نماد T_{60} نشان می‌دهند. به طور کلی زمان واخنش بهینه در بسامدهای میانی در فضاهای داخلی تصرف‌های آموزشی در جدول ۱ ارائه شده است [۴].

جدول ۱ زمان واخنش بهینه محیط‌های آموزشی.

نوع فضا	زمان واخنش بهینه (ثانیه)
کلاس درس نظری	۱٫۰
آزمایشگاه‌ها	۱٫۲
راهروها	۱٫۵

۲-۱. روش‌های مبتنی بر صوتیات هندسی

در صوتیات هندسی فرض ما این است که موج به صورت پرتوهای در راستای مستقیم پخش می‌شود. فرضیات مهم این روش این است که جبهه‌های موج را با پرتوهایی که عمود بر جبهه‌های موج هستند، می‌توان نمایش داد. این فرض در بسامدهای بالا جایی که طول موج صدا در مقایسه با ابعاد کلی فضا کوچک باشد، معتبر است اما این فرض در بسامدهای پایین دچار خطاهای در تقریب می‌شود [۳]. اصول صوتیات هندسی بسیار کلی است، و این فن‌های الگوسازی، به ویژه ردیابی پرتو و رادیوزیتی^۲، در زمینه‌های دیگری مانند اپتیک و صوتیات زیر آب استفاده می‌شوند. با این حال، این اصول بیش‌تر در گرافیک رایانه‌ای که در آن مساله، شبیه‌سازی رفتار نور در محیط است، استفاده می‌شود. در صوتیات هندسی فرض ما، بالا بودن بسامد امواج صوتی نسبت به جزئیات هندسه محیط و کوچک بودن طول موج است و می‌توان در این شرایط امواج صوتی را به صورت انتشار در راستای مستقیم فرض کرد در نتیجه این شباهت‌ها باعث شده همگرایی‌هایی با مباحث گرافیک رایانه‌ای مشاهده شود [۳].

۳-۱. بررسی تحقیقات پیشین

تلاش‌ها در این حوزه در سال‌های اخیر منجر به ارائه رویکردهایی جدید برای پیش‌بینی شبه‌سنج‌های صوتی

باعث می‌شود که این فرمول‌ها با مقادیر اندازه‌گیری‌شده اختلاف داشته باشند. از این‌رو به نظر می‌رسد که استفاده از معادلات کلاسیک یا هر رابطه دیگر که چنین فرضی را برای پیش‌بینی زمان واخنش داشته باشد، باید توأم با خطای زیادی باشد [۲]. در این راستا، یکی از مهم‌ترین موضوعات، استفاده از روش‌های جایگزین است که می‌تواند روابط پیچیده‌ای را بین متغیرهای مختلفی که بر زمان واخنش در اتاق‌های محصور تأثیر می‌گذارند، تشخیص دهد. رویکردهای جدیدتر پیش‌بینی شرایط صوتی محیط، روش‌های شبیه‌سازی مبتنی بر پرتو و روش‌های عنصر متناهی هستند که دقت بالاتری در مقایسه با فرمول‌های کلاسیک نشان می‌دهند. با این حال، استفاده از این روش‌ها نیاز به تخصص رایانه و صوتیات دارد و مستلزم محاسبات وقت‌گیر است. همچنین آن‌ها فقط در موارد خاص به کار می‌روند. زیرا الگوی هندسی ایجادشده برای یک اتاق خاص برای سایر اتاق‌ها کاربرد ندارد و عیب دیگری هم که در این رویکرد وجود دارد این است که هر یک از روش‌های شبیه‌سازی در یک بازه بسامدی خاص دقت بالایی دارند [۳].

استفاده از روش‌های مبتنی بر یادگیری ماشین می‌تواند به عنوان یک روش جایگزین برای پیش‌بینی زمان واخنش یک محیط ارائه شود. قابلیت این روش‌ها در تعیین شبه‌سنج‌های تأثیرگذار در پیش‌بینی زمان واخنش باعث می‌شود تا بتوان نقش عوامل مختلف را سنجید. عوامل محیطی به عنوان ورودی‌های این روش عمل می‌کنند و مقادیر زمان واخنش به عنوان خروجی به دست می‌آیند. در این بخش به ارائه مطالب کلی در مورد زمان واخنش و روش ردیابی پرتو^۱ پرداخته می‌شود و همچنین تحقیقات گذشته در حوزه موضوع پژوهش مورد بررسی قرار می‌گیرد.

۱-۱. زمان واخنش

یکی از شبه‌سنج‌های مهم در زمینه بررسی و تحلیل محیط‌های صوتی زمان واخنش می‌باشد. زمان واخنش شبه‌سنجی وابسته به بسامد است. مدت زمان لازم برای میراشدن صدا یا تحلیل رفتن صدا در یک فضای بسته را

² Radiosity

¹ Ray tracing

۲. مواد و روش‌ها

سامانه طراحی شده در این پژوهش که به منظور پیش‌بینی زمان واخشن مبتنی بر روی کردهای مبتنی بر یادگیری ماشین است، به چند بخش تقسیم می‌شود. ابتدا مجموعه دادگان با استفاده از روی کرد ردیابی پرتو و به کمک شبیه‌ساز اودئون جمع‌آوری می‌شود. سپس داده‌های خام جمع‌آوری شده به منظور آموزش و آزمایش الگو تقسیم‌بندی می‌شوند. از ۲۰ درصد مجموعه دادگان به منظور آزمایش شبکه استفاده می‌شود و بعد مجموعه آموزشی به عنوان ورودی شبکه عصبی استفاده می‌شود و شبه‌سنج‌های الگوی شبکه عصبی تنظیم می‌شوند و در نهایت مجموعه آزمایش روی الگوی آموزش داده شده قرار می‌گیرد و عملکرد الگو مورد ارزیابی قرار می‌گیرد. مجموعه دادگان شامل ۷۸۸ نمونه داده است و دارای ۱۴ ویژگی مختلف است. ورودی‌های اصلی این شبکه عصبی و اطلاعات آماری ورودی‌ها در جداول ۳ تا ۵ ارائه شده‌اند.

۲-۱. جمع‌آوری مجموعه دادگان

الگوریتم‌های یادگیری ماشین توسط داده‌ها تغذیه می‌شوند و عملکرد این روش‌ها برای هر عمل مشخصی با کیفیت و کمیت این داده‌ها محدود می‌شوند. برای مسائل مربوط به صوتیات اتاق، تعداد کمی مجموعه داده استاندارد، بزرگ و قابل دسترسی وجود دارد. در عوض، در بیش‌تر مطالعات، محققان، داده‌های خود را با توجه به نیازشان جمع‌آوری می‌کنند. اگر چه به اشتراک گذاشتن این داده‌ها امری معمول است، اما بیش‌تر اوقات به دلیل کوچک بودن، کمبود تنوع در ویژگی‌ها، قابلیت استفاده در پژوهش را ندارند. با این حال بیش‌تر مجموعه داده‌های موجود و قابل دسترسی، فاقد اطلاعات دقیق در مورد هندسه محیط و ضرایب جذب و اشیا قرار گرفته داخل اتاق هستند. بنابراین یک مجموعه داده جدید بنا به نیاز این پژوهش، تجزیه و تحلیل می‌شود. در این مقاله، با توجه به نظریه صوتیات هندسی مطرح شده در بخش (۱-۲)، با استفاده از شبیه‌ساز اودئون جمع‌آوری داده‌ست انجام می‌شود. در نتیجه به منظور داشتن یک مجموعه داده استاندارد و کافی برای استفاده در سامانه مبتنی بر یادگیری ماشین در فاز اول ابتدا چهار کلاس درس با ابعاد

شد. هدف این بود که هم عوامل ناشناخته در زمینه شبه‌سنج‌های صوتی کشف شود و هم اینکه بتوان روی کردی ارائه داد که به واقعیت نزدیک‌تر باشد. منبع [۵] به بررسی عملکرد شبیه‌سازی اودئون در اتاق‌های محصور می‌پردازد و عملکرد این شبیه‌سازی را با رگرسیون با مقادیر اندازه‌گیری شده مقایسه می‌کند. نووسویات [۲]، [۶] یک الگوریتم برای بهبود فرمول سابین ارائه داد و یک ضریب تصحیح وابسته به بسامد برای اصلاح این رابطه بیان کرد تا بتواند مشکل این فرمول را در فرض یکنواخت بودن میدان صوتی برطرف کند و مقادیر پیش‌بینی شده این فرمول سابین را به مقادیر اندازه‌گیری شده نزدیک‌تر کند. هم‌چنین فالکن و همکاران یک روش مبتنی بر یادگیری ماشین و با استفاده از شبکه عصبی پرسپترون چندلایه، برای تخمین شبه‌سنج صوتی در مسائل واقعیت مجازی طراحی کردند و نتایج حاصل از این کار را با مقدار اندازه‌گیری و نتایج حاصل از فرمول سابین و ایرینگ مقایسه کردند [۷]. این روش به دقتی تقریباً ۹۰ درصد برای بیش‌تر باندهای بسامدی رسیده است. وانگ و همکاران در مقاله‌ای به تخمین ابعاد اتاق با استفاده از شبکه عصبی کانولوشنی پرداختند [۸]. در این مقاله از تابع آرآی آر^۱ برای تولید داده‌ست استفاده شده است که این تابع توسط امانوئل هبت در سال ۲۰۱۰ معرفی شده بود [۹]. پاسخ ضربه‌های به‌دست آمده توسط تابع آرآی آر به‌صورت منحنی‌های دامنه-زمان به عنوان ورودی شبکه عصبی استفاده می‌شود، هم‌چنین خروجی شبکه عصبی استفاده شده در این مقاله سه نوریون دارد که برای طول، عرض و ارتفاع محیط به کار برده می‌شود. هم‌چنین در مقاله اوحدی و همکاران با استفاده از شبکه عصبی پرسپترون چندلایه برای تخمین زمان واخشن در محیط‌های نوفه‌ای صنعتی پرداخته شده است [۱۰]. در این کار ابتدا داده‌های حاصل از محیط‌های نوفه‌ای در کارگاه‌های گلدوزی جمع‌آوری می‌شوند و بعد با استفاده از یک شبکه عصبی ام‌ال‌پی^۲ به تخمین زمان واخشن می‌پردازد.

^۱ RIR; Room Impulse Responses

^۲ MLP

محدوده بسامد صحبت انسان، بین ۳۰۰ هرتز تا ۳ کیلوهرتز یا در بعضی مطالعات بین ۵۰۰ هرتز تا ۲ کیلوهرتز عنوان شده است [۱۱]. در این پژوهش تلاش شد که جمع‌آوری مجموعه دادگان تنها در یک تک بسامد صورت نگیرد در نتیجه با توجه به محدوده بسامد صحبت انسان و همچنین محدودیت در جمع‌آوری مجموعه دادگان، شبیه‌سازی‌ها فقط در دو بسامد ۵۰۰ هرتز و ۲۰۰۰ هرتز صورت گرفت. شبیه‌سازی در نرم‌افزار اودئون به این صورت است که، ابتدا الگوی طراحی شده وارد می‌شود سپس منابع صدا و گیرنده‌ها تعریف می‌شوند و بعد از این مرحله باید موادهای مورد نظر برای هر سطح انتخاب شوند. در مرحله بعد بنا به دقت مدنظر، تعداد پرتو انتخاب و تعیین می‌شود و بعد از آن بنا به نوع خروجی مدنظر، شبیه‌سازی انجام می‌شود. برای تمام شبیه‌سازی‌ها موقعیت استاد یا منبع صوتی در محلی که احتمال حضور استاد بیش‌تر است یعنی در کنار تخته آموزشی در نظر گرفته شد. به همین منظور در ارتفاع ۱۷۰ سانتی‌متری از کف محیط و به فاصله یک متری از دیوار و در وسط اتاق منبع صوتی قرار می‌گیرد. همچنین به تناسب با ابعاد محیط از تعداد گیرنده‌های متفاوتی در این محیط‌ها استفاده می‌شود. این گیرنده‌ها در مکان‌هایی که بیش‌ترین احتمال حضور دانشجویان را دارد انتخاب شده است. در این شبیه‌سازی‌ها از منبع تمام جهته^۱ استفاده شد، به دلیل اینکه معمولاً در بیش‌تر مواقع، اندازه‌گیری‌های استاندارد شبه‌سنج‌های صوتی با منابع تمام جهتی انجام می‌شود در نتیجه برای شبیه‌سازی هم از منابع تمام جهته استفاده شد، تا با نتایج اندازه‌گیری واقعی بتوان مقایسه نمود.

۲-۲. اطلاعات آماری مجموعه دادگان

در این پژوهش از ۷۸۸ نمونه داده برای آموزش و آزمایش، الگوریتم‌های یادگیری ماشین استفاده شد که اطلاعات آماری مجموعه دادگان در جدول ۳ آمده است. همچنین اطلاعات آماری ضرایب جذب کلی سطوح در بسامد ۵۰۰ هرتز در جدول ۴ ارائه گردیده است و اطلاعات آماری

مختلف و تعداد اشیاء مختلف، با استفاده از نرم‌افزار اسکچاپ طراحی شد. جدول ۲، اطلاعات کلاس‌های استفاده‌شده در طراحی را ارائه داده است.

جدول ۲ اطلاعات کلاس‌ها و تعداد گیرنده‌ها.

تعداد	حجم (مترمکعب)	ارتفاع (متر)	عرض (متر)	طول (متر)	
۴	۲۲۴	۲٫۸	۸	۱۰	اتاق ۱
۳	۱۷۵	۳٫۳	۷	۷٫۶	اتاق ۲
۳	۲۰۷	۳٫۳	۷	۹	اتاق ۳
۴	۲۷۲	۳٫۱	۸	۱۱	اتاق ۴

سپس شبیه‌سازی صوتی محیط‌ها به کمک نرم‌افزار اودئون انجام شد. برای هر کلاس ۵۶ ترکیب مختلف از موادهای برای سطوح مختلف در نظر گرفته شد، در انتخاب موادهای سعی شد که تنوع زیادی در ترکیب‌ها لحاظ شود، به گونه‌ای که حالات مختلف اتاق با مجموع ضریب جذب بالا تا مجموع ضریب جذب پایین، نسبت به موادهای موجود ایجاد شود و نتایج به‌دست‌آمده جامعیت خوبی داشته باشد. تخمین زمان واخنش در این پژوهش به این صورت است که، زمان واخنش در شبکه‌های عصبی استفاده‌شده به‌صورت تابعی از ویژگی‌های ورودی تعیین‌شده در جداول ۳ تا ۵ ارائه می‌شود.

$$RT_{60} = f(x_1, x_2, x_3, \dots, x_{14}) \quad (3)$$

که در این رابطه (۳)، RT_{60} ، زمان واخنش پیش‌بینی‌شده توسط شبکه عصبی است و x_1 تا x_{14} ویژگی‌های ورودی تعیین‌شده در مجموعه دادگان هستند که در جداول ۳ تا ۵، مشخصات و اطلاعات آماری این ویژگی‌ها آمده است. در شکل ۱ نمونه‌ی کلاس درس طراحی‌شده مشاهده می‌شود.



شکل ۱ نمونه کلاس درس طراحی‌شده.

¹ Omni directional

جدول ۳ اطلاعات آماری مجموعه دادگان.

نسبت موقعیت عرضی گیرنده به عرض اتاق	نسبت موقعیت طولی گیرنده به طول اتاق	حجم اتاق	ارتفاع	عرض	طول	میانگین
۰/۴۱۸	۰/۵۱۷	۲۲۴/۱۱	۳/۰۹۸	۷/۵۷	۵۵/۹	میانگین
۰/۰۳۴	۰/۲۳۷	۳۵/۱۲۸	۰/۲۰۷	۰/۴۹	۲۳/۱	انحراف معیار
۰/۴۱	۰/۵	۲۲۴	۳/۱	۸	۱۰	میان

جدول ۴ اطلاعات آماری جذب کلی سطوح در بسامد ۵۰۰ هرتز.

جذب پنجره	جذب دیوارها	جذب سقف	جذب کف	جذب میز و صندلی	میانگین
۱/۵۵	۱۳/۴۲	۲۱/۸۸	۱۸/۳۴	۱۵/۱۳۹	میانگین
۱/۲۱	۱۴/۳۰	۲۳/۷۱	۲۲/۷۲	۷/۳۹۱	انحراف معیار
۱/۱۷	۵/۸۴۷	۸/۸	۶/۴	۱۷/۰۵۶	میان

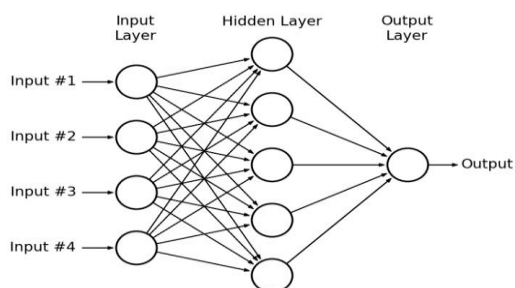
جدول ۵ اطلاعات آماری جذب کلی سطوح در بسامد ۲۰۰۰ هرتز.

جذب پنجره	جذب دیوارها	جذب سقف	جذب کف	جذب میز و صندلی	میانگین
۱/۵۵	۱۳/۴۲	۲۱/۸۸	۱۸/۳۴	۱۵/۱۳۹	میانگین
۱/۲۱	۱۴/۳۰	۲۳/۷۱	۲۲/۷۲	۷/۳۹۱	انحراف معیار
۱/۱۷	۵/۸۴۷	۸/۸	۶/۴	۱۷/۰۵۶	میان

for all features of input : $(I_1, I_2, \dots, I_n)$

$$y_i = \text{function} \left(\sum_{i=1}^n w_{ij} I_i + \text{Bias}_i \right) \quad (4)$$

در این رابطه، w وزن‌های شبکه عصبی و I ورودی‌های شبکه عصبی می‌باشند. در این پژوهش، در شبکه عصبی پرسپترون چندلایه از تابع فعال‌سازی تانژانت هیپربولیک، بهینه‌ساز آدام و با تعداد ۳۰ نورون در لایه پنهان طراحی شد و همچنین تعداد دور تکرار برابر ۲۰۰ در نظر گرفته شد.



شکل ۲ ساختار یک شبکه عصبی پرسپترون چندلایه [۱۲].

۴-۲. شبکه عصبی مبتنی بر توابع پایه شعاعی

یک شبکه عصبی مبتنی بر توابع پایه شعاعی را می‌توان به صورت یک شبکه سه لایه در نظر گرفت. در این شبکه

ضرایب جذب کلی سطوح در بسامد ۲۰۰۰ هرتز هم در جدول ۵ آمده است. جذب کلی سطوح به معنای حاصل ضرب سطح در ضریب جذب است.

۲-۳. شبکه عصبی پرسپترون چندلایه

شبکه عصبی پرسپترون چندلایه، دسته‌ای از شبکه‌های عصبی مصنوعی پیشخور محسوب می‌شود. در یک شبکه عصبی پرسپترون چندلایه، حداقل سه لایه از گره‌ها وجود خواهند داشت. در این شبکه عصبی، از خروجی‌های لایه اول (ورودی)، به عنوان ورودی‌های لایه بعدی (پنهان) استفاده می‌شود، این کار به همین شکل ادامه پیدا می‌کند، تا زمانی که، پس از تعداد مشخصی از لایه‌ها، خروجی‌های آخرین لایه پنهان به عنوان ورودی‌های لایه خروجی مورد استفاده قرار می‌گیرد. شکل ۲ ساختار یک شبکه عصبی پرسپترون چندلایه را نشان می‌دهد. الگوریتم پیشخور، فرآیندی را کنترل می‌کند که در نتیجه آن، ورودی‌های مسئله به یک خروجی تبدیل می‌شوند.

۲-۴-۱. روش خوشه‌بندی کی-میانگین

یک روش برای تعیین مرکز، استفاده از مراکز تصادفی است این روش ساده‌ترین و سریع‌ترین رویکرد است، ولی اگر توزیع داده‌ها به صورت یکسان نباشد این روش با خطا همراه خواهد بود، از این رو معمولاً برای آموزش شبه‌سنج‌های لایه میانی آربی‌اف از روش‌های خوشه‌بندی استفاده می‌شود [۱۳]. الگوریتم کی-میانگین^۳ شاید ساده‌ترین الگوریتم خوشه‌بندی باشد که در بسیاری از مواقع جزو بهترین الگوریتم‌های خوشه‌بندی نیز هست. این الگوریتم یکی از روش‌های یادگیری بدون نظارت^۴ است. این اصطلاح کی-میانگین ابتدا توسط مک کوین در سال ۱۹۶۷ مورد استفاده قرار گرفت. رایج‌ترین الگوریتم کی-میانگین با استفاده از یک تکرار شونده کار می‌کند [۵]. آن را با عنوان الگوریتم لوید نیز می‌شناسند.

در این روش ابتدا کی-میانگین یعنی $(\mu_1^{(0)}, \mu_2^{(0)}, \dots, \mu_k^{(0)})$ که نماینده خوشه‌ها هستند به صورت تصادفی مقداری اولیه می‌شوند. سپس، هر مشاهده یا شی به نزدیک‌ترین خوشه نسبت داده می‌شود. به این معنی که فاصله اقلیدسی هر مشاهده از مراکز، اندازه گرفته شده و سپس آن مشاهده‌ای عضو خوشه‌ای خواهد شد که کم‌ترین فاصله اقلیدسی را با مرکز آن خوشه دارد.

$$S_i^{(t)} = \left\{ x_p : \|x_p - \mu_i^{(t)}\| \leq \|x_p - \mu_j^{(t)}\| \right\} \quad \forall j, 1 \leq j \leq k \quad (7)$$

سپس میانگین خوشه‌های جدید محاسبه می‌شوند، این مرحله به تناوب تکرار خواهد شد. در نهایت میانگین‌های مرحله آخر (در زمان T) یعنی $(\mu_1^{(T)}, \mu_2^{(T)}, \dots, \mu_k^{(T)})$ خوشه‌ها را نمایندگی خواهند کرد.

در این پژوهش، در شبکه عصبی مبتنی بر توابع پایه شعاعی از ۳۰۰ کرنل گوسی، با انحراف معیار ۰.۵ استفاده شد. در شبکه عصبی مبتنی بر توابع پایه شعاعی از دو روش برای تعیین مراکز استفاده شد، که روش تصادفی و خوشه‌بندی کی-میانگین بود، که دقت بهتر با استفاده از روش خوشه‌بندی کی-میانگین به دست آمد. شبه‌سنج‌های تنظیمی الگو در الگوریتم کی-میانگین به این صورت

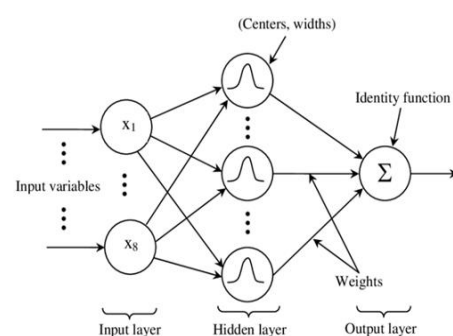
عصبی، خروجی به عنوان یک مساله تقریب تابع، به صورت ترکیب خطی از توابع گوسی به دست می‌آید. تبدیل فضای ورودی به فضای لایه پنهان، دارای رابطه غیرخطی است، ولی تبدیل فضای لایه پنهان به فضای خروجی دارای رابطه خطی است. توجیه ریاضیاتی انتخاب این توابع تبدیل توسط، کاور^۱ در سال ۱۹۶۵، بدین گونه بیان شده که برای جداسازی الگوهای پیچیده بین داده‌ها، فضای ورودی را به فضایی با بعد بالا منتقل کرده تا بتوان جداسازی راحت‌تری داشت. یعنی فضای ورودی با یک نگاشت غیرخطی به فضای ویژگی تبدیل می‌شود، و ترکیب خطی این‌ها یک خروجی می‌دهد [۱۳]. ساختار شبکه عصبی مبتنی بر توابع پایه شعاعی به صورت رابطه (۵) است. در این رابطه، x بردار ورودی، g تابع پایه شعاعی، w وزن‌های لایه خروجی و y خروجی شبکه می‌باشند. خروجی شبکه از رابطه (۵) به دست می‌آید و c_i شبه‌سنج مرکز توابع مربوط به گره i می‌باشد.

$$y = \sum_i^k w_i g_i(x) = \sum_i^k w_i g_i(x - c_i) \quad (5)$$

تابع گوسی معمولاً به صورت رابطه (۵) بیان می‌شود:

$$g_i(x) = e^{-\frac{x - c_i}{\sigma^2}} \quad (6)$$

با توجه به این ساختمان، آموزش شبکه آربی‌اف^۲ عبارت است از یافتن مقادیر مناسبی از مراکز و گسترده‌گی گوسی‌ها و همچنین وزن‌های لایه آخر. شکل ۳ ساختار یک شبکه عصبی مبتنی بر توابع پایه شعاعی را نشان می‌دهد.



شکل ۳ ساختار شبکه عصبی مبتنی بر توابع پایه شعاعی [۱۳].

^۳ k-means clustering

^۴ Unsupervised learning

^۱ cover

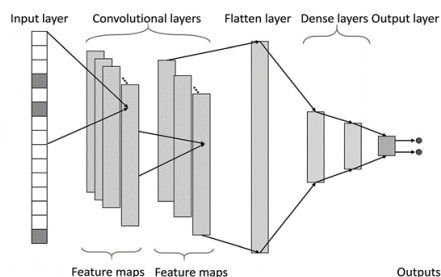
^۲ RBF

بعدی به لایه شبکه تمام متصل ارسال می‌شود. لایه‌های تمام متصل معمولاً در انتهای معماری‌های شبکه کانولوشن استفاده می‌شوند [۱۵]. ساختار الگوی سی‌ان‌ان-۱^۱ و تنظیم شبه‌سنگ‌ها در این مقاله از طریق چندین آزمایش به‌دست آمده است. ساختار این شبکه عصبی کانولوشن یک-بعدی، شامل شش لایه کانولوشن، دو لایه پولینگ و دو لایه تمام متصل است. جدول ۶ شبه‌سنگ‌های معماری شبکه سی‌ان‌ان-۱^۱ را نشان می‌دهد. در لایه اول از ۸ صافی، در لایه دوم از ۱۶ صافی، در لایه سوم از ۳۲ صافی، در لایه چهارم از ۶۴ صافی، در لایه پنجم از ۱۲۸ صافی و در لایه ششم از ۲۵۶ صافی با اندازه ۳ استفاده شد. صافی‌ها فقط آرایه‌ای از اعداد هستند که در فرآیند آموزش برورسانی می‌شوند. به حرکت صافی استفاده‌شده روی آرایه ورودی پیچش یا کانالو می‌گویند. با استفاده از این صافی‌ها روی داده‌های ورودی پردازش می‌شود. هدف از صافی‌های کانولوشن، استخراج ویژگی‌های سطح بالا از ورودی است. اعداد موجود در صافی‌ها طی یک فرآیند آموزش، از طریق الگوریتم پس انتشار خطا به‌دست می‌آیند [۱۵]. هم‌چنین برای شبه‌سنگ پدینگ^۲ در این پژوهش مقدار یکسان^۳ در نظر گرفته می‌شود. از آنجایی که قسمت‌هایی از آرایه که در حاشیه قرار می‌گیرد و نمی‌توان با ابعاد باقی‌مانده عملیات کانولوشن را انجام داد به همین خاطر، جهت انجام عملیات به نواحی مذکور صفر اضافه می‌شود. هم‌چنین مشابه با سایر شبکه‌های عصبی، شبکه عصبی کانولوشن هم از تابع تحریک غیرخطی بعد از لایه کانولوشن استفاده می‌کند. از مهم‌ترین توابع فعال‌ساز استفاده‌شده می‌توان به تابع سافت مکس، تابع رلو، تابع الو اشاره کرد [۱۵]. در این پژوهش از تابع فعال‌ساز رلو، بعد از لایه‌های کانولوشن استفاده شد. هم‌چنین در این پژوهش به‌منظور حداقل کردن تابع هزینه و برورسانی شبه‌سنگ‌های شبکه، از الگوریتم بهینه‌ساز آدام استفاده شد. هم‌چنین در این معماری از دو لایه پولینگ استفاده شد. لایه پولینگ معمولاً بعد از یک لایه کانولوشن قرار

بودند که تعداد خوشه‌ها برابر با ۳۰۰ در نظر گرفته شدند هم‌چنین تعداد فرآیند تکرار برای رسیدن به نتیجه مطلوب در این الگوریتم برابر با ۳۰۰ در نظر گرفته شد. هم‌چنین برای به‌دست آوردن شبه‌سنگ‌های وزن، از روش حل معادله نرمال استفاده شد.

۲-۵. شبکه عصبی کانولوشن

شبکه عصبی کانولوشنی از انواع شبکه عصبی و یکی از الگوهای پر استفاده امروزی، در مباحث یادگیری ماشین است. دلیل نام‌گذاری این نوع شبکه عصبی، وجود لایه کانولوشن یا پیچشی، در مرکز شبکه است. این لایه، عملیاتی به نام کانولوشن را انجام می‌دهد. شبکه عصبی کانولوشنی از پرسپترون‌های چندلایه و یک یا چندلایه کانولوشن تشکیل شده است. شکل ۴ یک الگوی ساده از شبکه عصبی کانولوشن یک-بعدی است. به دلیل استفاده از داده‌های یک-بعدی از شبکه عصبی کانولوشن یک-بعدی استفاده شد. در این گونه از شبکه عصبی، صافی‌ها (فیلترها) فقط در یک جهت حرکت می‌کنند در صورتی که در شبکه عصبی کانولوشن دو-بعدی و سه-بعدی که عمدتاً برای کار روی تصاویر و ویدئو است، این صافی‌ها در دو و سه جهت حرکت می‌کنند.



شکل ۴ ساختار شبکه عصبی کانولوشن یک-بعدی [۱۴].

لایه اول در این شبکه عصبی، همیشه یک لایه کانولوشن است، ورودی این لایه یک آرایه از اعداد است. شبکه‌های کانولوشنی معمولاً به یک لایه کانولوشنی محدود نمی‌شوند و لایه‌های بیش‌تری دارند، در هریک از این لایه‌ها ویژگی خاصی تشخیص داده می‌شود. معمولاً، لایه‌های ابتدایی برای استخراج ویژگی‌های سطح پایین و لایه‌های بعدی برای استخراج ویژگی‌های سطح بالا به‌کار برده می‌شود. خروجی این لایه بعد از تبدیل به بردار یک-

¹ CNN-1d

² Padding

³ Same

متغیرهای مستقل استفاده می‌شود هیچ خطایی وجود ندارد که این بهترین حالت ممکن است. اگر $R^2 = 0$ باشد یعنی استفاده از متغیرهای مستقل هیچ تاثیری بر برآورد خط رگرسیونی ندارد. این ضریب بین صفر تا ۱ تغییر می‌کند و مقدار بهینه آن ۱ است و این حالت مربوط به زمانی است که در آن مقادیر شبیه‌سازی شده دقیقاً مشابه مقادیر مشاهداتی باشند.

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_{imeas} - x_{ipred})^2}{\sum_{i=1}^n (x_{imeas} - \bar{x})^2} \quad (8)$$

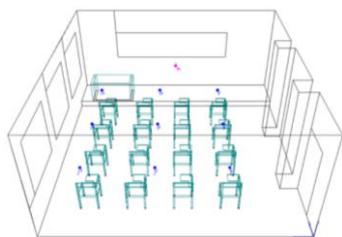
روی کرد دیگری که برای ارزیابی الگو استفاده شد، میانگین مربعات خطا بود. این روی کرد روشی برای برآورد میزان خطاست که در واقع تفاوت بین مقادیر اندازه‌گیری و آن چه تخمین زده شده، می‌باشد. این شاخص که مقداری همواره نامنفی دارد، هر چقدر مقدار آن به صفر نزدیک‌تر باشد، نشان‌دهنده میزان کم‌تر خطاست. مقدار این شاخص به صورت رابطه (۹) بیان می‌شود:

$$mse = \frac{1}{n} \sum_{i=1}^n [(x_{imeas} - x_{ipred})^2] \quad (9)$$

که در این رابطه x_{imeas} و x_{ipred} به ترتیب برابر با تعداد متغیر اندازه‌گیری شده، مقدار متغیر پیش‌بینی شده و مقدار متغیر اندازه‌گیری شده می‌باشند.

۳. نتایج

در این مرحله به منظور بررسی و اطمینان از نتایج نرم‌افزار شبیه‌ساز اودئون، در دو بسامد ۵۰۰ هرتز و ۲۰۰۰ هرتز، از نتایج اندازه‌گیری واقعی استفاده شد، تا اطمینان حاصل شود که مجموعه دادگان شبیه‌سازی شده، جایگزین مناسبی برای اندازه‌گیری واقعی هستند. به این منظور کلاس شماره ۲۰۲ دانشگاه صدا و سیما مورد بررسی قرار گرفت. شکل ۵ شبیه‌سازی اتاق شماره ۲۰۲ را نشان می‌دهد.



شکل ۵ شبیه‌سازی کلاس ۲۰۲ دانشگاه صدا و سیما در نرم‌افزار اودئون.

می‌گیرد و از آن برای کاهش اندازه نقشه‌های ویژگی و شبه‌سنتج‌های شبکه می‌توان استفاده کرد. آن چه در این لایه اتفاق می‌افتد، کاهش ابعاد است. لایه پولینگ شبه‌سنتج قابلیت آموزش ندارد و صرفاً یک نمونه‌برداری ساده و موثر انجام می‌دهد.

جدول ۶ معماری شبکه عصبی کانولوشن.

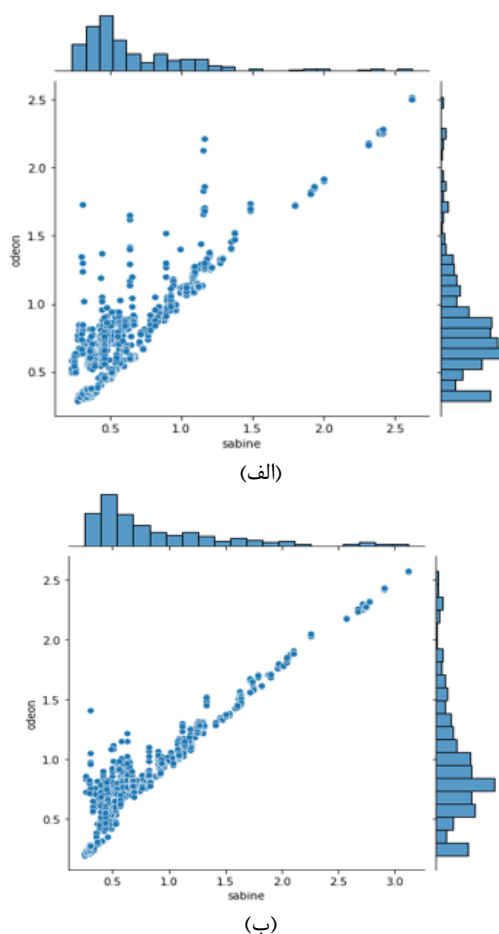
فعال‌سازی	شبه‌سنتج	لایه
ReLU	Filter\kernel۳	Conv\1D
ReLU	Filter\۱۶/kernel۳	Conv\1D
ReLU	Filter\۳۲/kernel۳	Conv\1D Maxpooling
ReLU	Filter\۶۴/kernel۳	Conv\1D
ReLU	Filter\۱۲۸/kernel۳	Conv\1D
ReLU	Filter\۲۵۶/kernel۳	Conv\1D AveragePooling
ReLU	۲۰	Dense
Linear	۱	Dense

پولینگ عملکردی شبیه کانولوشن دارد و یک پنجره روی خروجی حرکت می‌کند. رایج‌ترین روی کرد پولینگ، میانگین‌گیری و انتخاب بیشینه مقادیر است. در روی کرد اول، میانگین مقادیر موجود در هر کدام از پنجره‌ها محاسبه می‌شود و در روش دوم، بیش‌ترین مقدار موجود در این پنجره محاسبه شده و به عنوان مقدار در نقشه ویژگی جدید انتخاب می‌شود [۱۵].

۲-۶. معیارهای ارزیابی

ضریب تعیین که با R^2 نشان داده می‌شود قدرت توضیح‌دهندگی الگو را نشان می‌دهد [۷، ۱۰]. ضریب تعیین نشان می‌دهد که چند درصد از تغییرات متغیر وابسته توسط متغیرهای مستقل توضیح داده می‌شود. در واقع این ضریب بیانگر میزان احتمال هم‌بستگی میان دو دسته داده در آینده می‌باشد. این ضریب در واقع نتایج تقریبی شبه‌سنتج مورد نظر در آینده را براساس الگوی ریاضی تعریف شده که منطبق بر داده‌های موجود است، بیان می‌دارد. ضریب تعیین، معیاری است از این که خط رگرسیون، چقدر خوب خوانده‌ها را معرفی می‌کند. اگر خط رگرسیون از تمام نقاط بگذرد توانائی معرفی همه متغیرها را دارد و هر چه از نقاط دورتر باشد نشان‌دهنده توانایی کم‌تر است. اگر $R^2 = 1$ باشد یعنی زمانی که از

شبیه‌سازی، در این مرحله با استفاده از مجموعه دادگان ساخته‌شده حاصل از شبیه‌سازی، ابتدا نتایج فرمول سابین استخراج می‌شوند و این نتایج با نتایج شبیه‌سازی اودئون مقایسه می‌گردند. در بررسی‌های صورت‌گرفته بین نتایج شبیه‌سازی اودئون و نتایج حاصل از فرمول سابین، از نمودار توزیع مشترک بین سابین و اودئون استفاده شد. در بسامد ۵۰۰ هرتز، همان‌طور که از شکل ۶ مشخص است، ارتباط خطی بین نتایج حاصل از سابین و نتایج شبیه‌سازی وجود ندارد.



شکل ۶ نمودار توزیع مشترک نتایج زمان واخنش محاسبه‌شده با اودئون و سابین در بسامد (الف) ۵۰۰ هرتز و (ب) ۲۰۰۰ هرتز.

هم‌چنین برای بسامد ۲۰۰۰ هرتز، همان‌طور که مشخص است باز هم ارتباط کاملاً خطی بین نتایج حاصل از سابین و نتایج شبیه‌سازی وجود ندارد، ولی ارتباط خطی‌تری نسبت به بسامد ۵۰۰ هرتز وجود دارد. علت این است که در بسامدهای نزدیک‌تر به بسامد شرودر در یک محیط، به

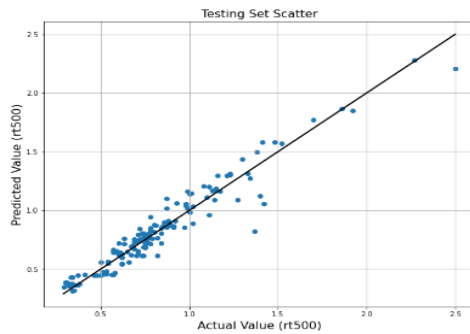
این کلاس تقریباً مستطیل شکل است و به طول ۶ متر، عرض ۶/۹ و ارتفاع ۳/۳ متر می‌باشد. در نتیجه نتایج زمان واخنش این کلاس با شبیه‌سازی حاصل از اودئون مقایسه می‌شود. هم‌چنین با توجه به ثبت نه اندازه‌گیری، از نه گیرنده با توجه به موقعیت‌های مختلفی که احتمال حضور دانش آموزان وجود دارد، استفاده شد.

در این اندازه‌گیری از منابع تمام‌جهته استفاده شده است. نتایج به‌دست‌آمده از شبیه‌سازی در جدول ۷، نشان‌دهنده دقت قابل قبول شبیه‌سازی است. برای بررسی بیش‌تر می‌توان گفت که نتایج جدول ۷ بیانگر این نکته هستند که دقت شبیه‌سازی با اودئون در بسامدهای بالاتر، بیش‌تر است، زیرا در محدوده بسامد بالا به‌علت اینکه فرضیات صوتیات هندسی به خوبی رعایت می‌شود در نتیجه شبیه‌سازی‌ها خطای کم‌تری دارند. برای بررسی نتایج در محدوده بسامد پایین، با استفاده از مفهوم بسامد شرودر تحلیل انجام می‌شود. طبق گفته شرودر [۱۶] در محدوده بسامدی بالاتر از بسامد شرودر، رفتار صوتی به صورت اشعه‌ای خواهد بود و در محدوده پایین‌تر از بسامد شرودر، امواج صوتی به صورت موجی رفتار خواهد کرد. در نتیجه از آن‌جایی که بسامد شرودر برای کلاس ۲۰۲ عددی تقریباً برابر با ۱۸۰ هرتز است. و با توجه به ضریب چهار در نظر گرفته‌شده برای انتقال به ناحیه با حالت اشعه‌ای [۱۶]، شبیه‌سازی‌ها در بسامد ۵۰۰ هرتز با کمی خطای بیش‌تر همراه است.

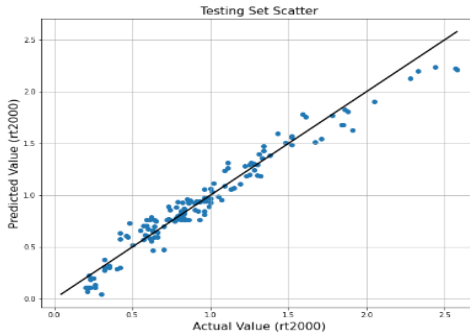
جدول ۷ نتایج زمان واخنش اندازه‌گیری‌شده و شبیه‌سازی‌شده.

بسامد	بسامد	
۵۰۰ هرتز	۲۰۰۰ هرتز	
۱/۴۸ (ثانیه)	۱/۱۸ (ثانیه)	اندازه‌گیری واقعی
۱/۵۴ (ثانیه)	۱/۲۰ (ثانیه)	نتایج حاصل از شبیه‌سازی اودئون
۱/۶۷ (ثانیه)	۱/۲۳ (ثانیه)	نتایج سابین

همان‌طور که از نتایج جدول ۷ مشخص است، شبیه‌ساز اودئون می‌تواند با دقت بالایی زمان واخنش را محاسبه کند. به همین جهت از نتایج شبیه‌ساز اودئون برای آموزش و آزمون الگوی مبتنی بر یادگیری ماشین استفاده می‌شود. هم‌چنین به‌منظور بررسی بیش‌تر نتایج

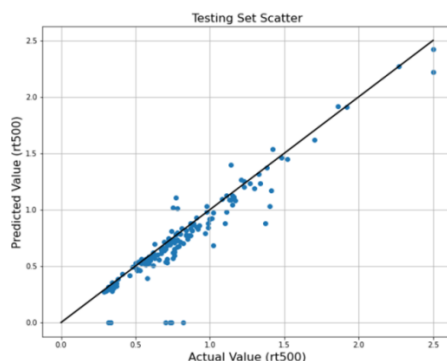


(الف)

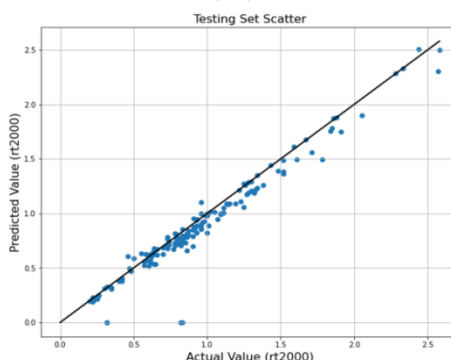


(ب)

شکل ۷ نتایج زمان واخشن پیش‌بینی‌شده با شبکه عصبی پرسپترون چندلایه در دو بسامد، (الف) ۵۰۰ هرتز و (ب) بسامد ۲۰۰۰ هرتز.



(الف)



(ب)

شکل ۸ نتایج زمان واخشن پیش‌بینی‌شده با شبکه عصبی مبتنی بر توابع پایه شعاعی در دو بسامد (الف) ۵۰۰ هرتز و (ب) بسامد ۲۰۰۰ هرتز.

علت وجود امواج ایستاده، وابستگی به محل قرارگیری گیرنده بیش‌تر است و به همین علت در بسامدهای پایین‌تر، فرمول سابین دقت کم‌تری دارد. همچنین نشان داده شد که فرض یکنواخت بودن میدان صوتی محیط، باعث ایجاد خطا در پیش‌بینی می‌شود و فرمول کلاسیک تخمین زمان واخشن نمی‌تواند به خوبی رفتار صوتی یک محیط را پیش‌بینی کند. بنابراین، مشهود است که به منظور داشتن دقت بیش‌تر در پیش‌بینی زمان واخشن باید از شبه‌سنج‌های بیش‌تری نسبت به فرمول سابین استفاده شود. در نتیجه لزوم استفاده از رویکردهای جایگزین برای پیش‌بینی زمان واخشن ثابت می‌شود.

۳-۱. نتایج شبکه عصبی

در این مرحله نتایج شبکه عصبی استفاده‌شده به منظور پیش‌بینی زمان واخشن مورد بررسی قرار می‌گیرد. در این پژوهش از مجموعه دادگان تولیدشده با رویکرد صوتیات هندسی و به کمک نرم‌افزار اودئون استفاده شد. شبه‌سنج‌های تنظیمی شبکه‌های عصبی پرسپترون چندلایه و شبکه عصبی مبتنی بر توابع پایه شعاعی در بخش (۲-۳) و (۲-۴) آمده‌اند، همچنین جزئیات معماری شبکه عصبی کانولوشن یک-بعدی، در این پژوهش در بخش (۲-۵) آمده است. همان‌طور که در جدول ۸، نتایج شبکه عصبی آمده است، با توجه به معیارهای ارزیابی مطرح‌شده در بخش (۲-۶)، این سامانه طراحی شده کارایی مناسبی برای پیش‌بینی زمان واخشن در دو بسامد ۵۰۰ و ۲۰۰۰ هرتز ثبت کرده است. همان‌طور که در جدول ۸ مشخص است در این الگو ضریب تعیین به‌دست‌آمده در بسامد ۲۰۰۰ هرتز بیش‌تر از بسامد ۵۰۰

جدول ۸ نتایج شبکه‌های عصبی استفاده‌شده در فاز آموزش و آزمایش

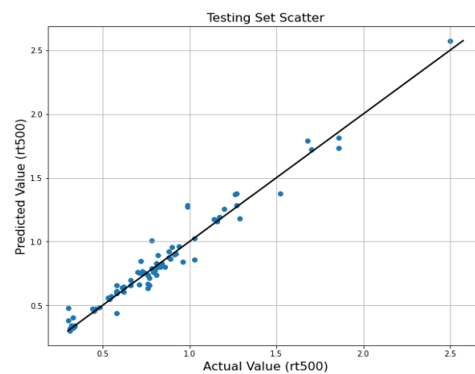
	بسامد ۵۰۰ هرتز		بسامد ۲۰۰۰ هرتز	
	R^2	آرام‌سازی	R^2	آرام‌سازی
ایم‌ال‌پی آموزش	۰/۹۲	۰/۱۰۵	۰/۹۴	۰/۱۱۲
آر‌بی‌اف آموزش	۰/۹۸	۰/۰۴۴	۰/۹۹	۰/۰۴۳
سی‌ان‌ان آموزش	۰/۹۹	۰/۰۶۳	۰/۹۸	۰/۰۳۱
ایم‌ال‌پی آزمایش	۰/۹۳	۰/۰۹۹	۰/۹۵	۰/۱۰۵
آر‌بی‌اف آزمایش	۰/۸۲	۰/۰۹۰	۰/۸۹	۰/۰۷۲
سی‌ان‌ان آزمایش	۰/۹۴	۰/۰۷۷	۰/۹۶	۰/۰۸۳

۴. بحث

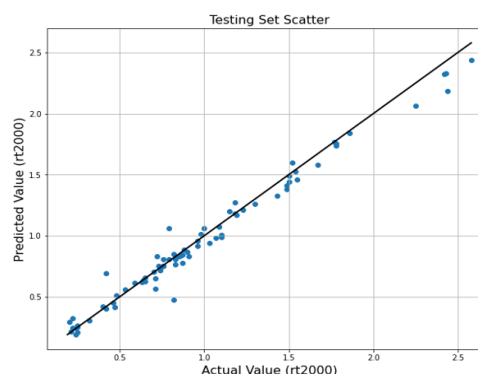
با توجه به اهمیت پیش‌بینی زمان واخنش به‌عنوان یکی از مهم‌ترین شبه‌سنج‌های صوتی در طراحی محیط‌های آموزشی، هدف از این پژوهش، پیش‌بینی زمان واخنش در نظر گرفته شد. عدم کارایی مناسب فرمول‌های کلاسیک مانند سابقین، باعث شد که در این مقاله به بررسی استفاده از روش‌های یادگیری ماشین به‌عنوان یک روش جدید و ابتکاری برای پیش‌بینی زمان واخنش محیط پرداخته شود، که به‌عنوان یک روش جایگزین برای روش‌های کلاسیک در نظر گرفته می‌شود. در نتیجه، در این مقاله یک سامانه برای تخمین زمان واخنش با استفاده از الگوریتم‌های یادگیری ماشین و به‌صورت یادگیری نظارت شده، ارائه گردیده است. به‌طوری که با استفاده از مجموعه دادگان ثبت‌شده بتوان یک روش جایگزین برای فرمول سابقین جهت پیش‌بینی زمان واخنش طراحی کرد. به این منظور ابتدا، به جمع‌آوری مجموعه دادگان پرداخته شد و سپس الگو، با در نظر گرفتن سه شبکه عصبی مختلف بررسی شد، و با تنظیمات مختلف روی شبه‌سنج‌های تنظیمی، نتایج با دقت مناسب ذخیره شدند. در این مطالعه، ابعاد و حجم محیط و مقادیر ضرایب جذب تفکیک‌شده سطوح و موقعیت قرارگیری گیرنده، به‌عنوان ورودی برای این الگو پیشنهادی تعیین شدند و زمان واخنش به‌عنوان خروجی این سامانه، در نظر گرفته شد. همان‌طور که از نتایج ثبت‌شده در جدول ۸ مشخص است شبکه عصبی کانولوشن یک-بعدی و شبکه عصبی پرسپترون چندلایه به ترتیب نتایج بهتری را نشان می‌دهند. برای بررسی با کارهای دیگر می‌توان به مواردی اشاره کرد، در منبع [۸] تنها از پاسخ ضربه به‌عنوان ورودی شبکه عصبی استفاده شد در صورتی که در مقاله حاضر از ۱۴ ویژگی مختلف برای ورودی شبکه عصبی استفاده شد، همچنین در این مقاله، از سه شبکه عصبی مختلف استفاده شد.

در این مقاله میانگین ضریب تعیین ثبت‌شده به‌عنوان یک شبه‌سنج ارزیابی الگوی پیشنهادی، نسبت به میانگین ضریب تعیین ثبت‌شده در مقاله فالکن [۷] نتیجه مطلوب‌تری را نشان می‌دهد. در منبع [۷]، با مجموعه دادگانی که شامل ۸۳۲ نمونه اندازه‌گیری‌شده واقعی در

هرتز است، که تا حد زیادی به این علت است که در مجموعه دادگان موجود، در بسامد پایین‌تر، زمان واخنش ثبت‌شده، تغییرات بیش‌تری نسبت به زمان واخنش در بسامد ۲۰۰۰ هرتز دارند. همچنین به‌منظور بررسی بیش‌تر در این پژوهش، نمودار نتایج شبکه عصبی در شکل‌های ۷ تا ۹ رسم شد. نتایج نشان می‌دهند که به ترتیب شبکه عصبی کانولوشن و شبکه عصبی پرسپترون چندلایه، پیش‌بینی مناسب‌تری با توجه به معیارهای ارزیابی از خود نشان می‌دهند. همان‌طور که از شکل‌های ۷ تا ۹ مشخص است، سامانه طراحی شده در این پژوهش با در نظر گرفتن ویژگی‌های لحاظ شده در مجموعه دادگان، توانست یک الگو برای پیش‌بینی زمان واخنش ارائه دهد، و می‌تواند زمان واخنش را برای دادگان آزمایش به میزان خوبی پیش‌بینی کند و به‌طور خاص همان‌طور که در شکل ۹ مشخص است پراکندگی نتایج در شبکه عصبی کانولوشن نسبت به مقدار واقعی زمان واخنش در مقایسه با سایر شبکه‌های عصبی کمتر است.



(الف)



(ب)

شکل ۹ نتایج زمان واخنش پیش‌بینی‌شده با شبکه عصبی کانولوشن در دو بسامد (الف) ۵۰۰ هرتز و (ب) ۲۰۰۰ هرتز.

فهرست منابع

- [1] N. Prodi, C. Visentin, A. Farnetani, "Intelligibility, listening difficulty and listening efficiency in auralized classrooms," *Journal of the Acoustical Society of America*, vol. 128, no. 1, pp. 172–181, 2010.
- [2] A. Nowoświat, M. Olechowska, J. Jusarek, "Prediction of reverberation time using the residual minimization method," *Applied Acoustics*, vol. 106, pp. 42–50, 2016.
- [3] L. Savioja, U.P. Svensson, "Overview of geometrical room acoustic modeling techniques," *Journal of the Acoustical Society of America*, vol. 138, no. 2, pp. 708–730, 2015.
- [4] Insulation and sound regulation-The eighteenth topic of Iran's national building regulations, 2017, (In Persian).
- [5] A. Nowoświat, M. Olechowska, "Experimental validation of the model of reverberation time prediction in a room," *Buildings*, vol. 12, no. 3, pp. 347, 2022.
- [6] A. Nowoświat, M. Olechowska, "Estimation of reverberation time in classrooms using the residual minimization method," *Arch Acoustics*, vol. 42, no. 4, pp. 609–617, 2017.
- [7] R. Falcón Pérez, G. Götz, V. Pulkki, "Machine-learning-based estimation of reverberation time using room geometry for room effect rendering," *International Congress on Acoustics*, pp. 7258–7265, 2019.
- [8] W. Yu, W.B. Kleijn, "Room geometry estimation from room impulse responses using convolutional neural networks," *Electrical Engineering and systems Science*, arXiv, vol. 1, 2019.
- [9] E. Habets, "Room impulse response generator," *Tutorial*, pp. 1–21, 2010.
- [10] M. Aliabadi, R. Golmohammadi, A. Ohadi, Z. Mansoorizadeh, H. Khotanlou, M.S. Sarrafzadeh, "Development of an empirical acoustic model for predicting reverberation time in typical industrial workrooms using artificial neural networks," *Acta Acustica united with Acustica*, vol. 100, no. 6, pp. 1090–1097, 2014.
- [11] K.B. Bhangale, K. Mohanaprasad, "A review on speech processing using machine learning paradigm," *International Journal of Speech Technology*, 2021.
- [12] A.H. Fath, F. Madanifar, M. Abbasi, "Implementation of multilayer perceptron (MLP) and radial basis function (RBF) neural networks to predict solution gas-oil ratio of crude oil systems," *Petroleum*, vol. 6, no. 1, pp. 80-91, 2020.

یک کلاس درس واحد بود به طراحی یک سامانه پیش‌بینی زمان واخنش پرداختند. نکته بعدی این است که در داده‌ست مقاله فالکن فقط در یک اتاق و با وجود عنصرهای مختلف در اتاق ثبت شده است. در صورتی که در این مقاله از چهار کلاس درس مختلف در ابعاد و حجم‌های مختلف استفاده شد که این یک مزیت عمده نسبت به کار فالکن محسوب می‌شود. در مقاله فالکن حجم و ابعاد اتاق به عنوان ورودی برای شبکه عصبی در نظر گرفته نشده، هم‌چنین تغییراتی در نظر گرفته نشده است. در مقاله [۱۰]، با استفاده از شبکه عصبی پرسپترون چندلایه و به کمک مجموعه داده‌گان اندازه‌گیری‌شده در کارگاه‌ها، به این صورت که از میانگین زمان واخنش اندازه‌گیری‌شده در باندهای بسامدی به عنوان شبه‌سنج هدف استفاده شده بود به ضریب تعیین ۷۷ درصد رسیدند. در این مقاله در مقایسه با مقاله اوحدی [۱۰]، کارایی بالاتری را ثبت کرده است.

۵. نتیجه‌گیری

در این مقاله، یک سامانه برای پیش‌بینی زمان واخنش با استفاده از رویکرد یادگیری نظارت‌شده و به کمک شبکه عصبی، بیان شد مهم‌ترین مشخصات این تحقیق عبارتند از؛ اهمیت توسعه رویکردهای جایگزین فرمول‌های کلاسیک برای پیش‌بینی زمان واخنش، استفاده از رویکردهای صوتیات هندسی در تولید مجموعه داده‌گان و استفاده از شبکه‌های عصبی برای پیش‌بینی شبه‌سنج زمان واخنش و مقایسه آن‌ها.

۶. تقدیر و تشکر

این مقاله از توسعه پایان‌نامه کارشناسی‌ارشد با عنوان «تخمین زمان واخنش کلاس درس مبتنی بر روش یادگیری ماشین و با استفاده از مجموعه داده تولیدشده با روش‌های صوتی هندسی» به‌دست آمده که با راهنمایی دکتر معصومه شفیعیان از دانشگاه صدا و سیما انجام گرفته است.

۷. تضاد منافع

این مطالعه تضاد منافع ندارد.

- [13] Y. Wu, H. Wang, B. Zhang, K.L. Du, "Using radial basis function networks for function approximation and classification," International Scholarly Research Notices, 2012.
- [14] Q. Zhao, Q. Mao, Z. Zhao, T. Dou, Z. Wang, X. Cui, Y. Liu, X. Fan, "Prediction of plant-derived xenomiRs from plant miRNA sequences using random forest and one-dimensional convolutional neural network models," BMC Genomics, vol. 19, no. 1, pp. 1-13, 2018.
- [15] N. Aloysius, M. Geetha, "A review on deep convolutional neural networks," International Conference on Communication and Signal Processing (ICCSP), pp. 0588-0592, 2017.
- [16] S.M. Khezri, P.J. Shalkouhi, "The schroeder frequency of furnished and unfurnished spaces," Romanian Journal of Acoustics and Vibration, vol. 9, no. 2, pp. 113, 2012.