



Improving the performance of the convolutional neural network using incremental weight loss function to deal with class imbalanced data

N. Mahmoodi, H. Shirazi^{ID}*, M. Fakhredanesh, K. Dadashtabar

* Malek Ashtar University of Technology , Tehran, Iran.

(Received: 2023/08/30, Revised: 2023/11/12, Accepted: 2024/01/02, Published: 2024/12/27)

DOR: <https://dorl.net/dor/20.1001.1.23224347.1402.11.4.2.9>

ABSTRACT

Class-imbalanced datasets are common in many real-world domains, such as health, banking, and security. Machine learning researchers have recently focused on the classification of such datasets, where the costs of different types of misclassifications are unequal, the classes have different prior probabilities, or both. The performance of most standard classifier learning algorithms is significantly affected by class imbalance, where the algorithms are often biased toward the majority class instances despite recent advances in deep learning. However, there is very little empirical work on deep learning with class imbalance. To address this issue, we propose an incremental weighted cross entropy loss function. The proposed method involves gradually increasing the weight of the minority class as the training progresses, until it reaches the specified amount at the end of the training. Through experiments, we demonstrate the convergence and efficiency of the proposed method. The results of experiments on three datasets, including artificial datasets, human activity recognition dataset, and CIFAR-10, demonstrate the convergence and performance of the proposed method. The proposed method is compared with decision tree-based AdaBoost, Cross Entropy-based convolutional neural network, weighted Cross Entropy -based CNN, SMOTE method, and ensemble CNNs method. With accuracy gains of 94.6%, 92.92%, and 69.23% on the three datasets (CIFAR-10 with 5% imbalance rate), the proposed method outperformed the other methods. Additionally, the accuracy on the artificial dataset was 17.77% higher than the traditional decision tree-based AdaBoost method.

Keywords: class-imbalanced dataset, Convolutional Neural Network, loss function, cross-entropy .

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Publisher: Imam Hussein University

 Authors



* Corresponding Author Email: shirazi@mut.ac.ir

علمی - پژوهشی

بهبود کارایی شبکه عصبی کانولووشنال با استفاده از تابع ضرر وزن دار افزایشی برای مقابله با نامتوازی دسته‌ای

نسیبه محمودی^۱، حسین شیرازی^{۲*}، محمد فخر دانش^۳، کوروش داداش تبار^۴

۱، ۲، ۳- دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی مالک اشتر

(دریافت: ۱۴۰۲/۰۴/۱۳، بازنگری: ۱۴۰۲/۰۹/۲۷، پذیرش: ۱۴۰۲/۱۰/۱۲، انتشار: ۱۴۰۲/۱۰/۲۲)

DOR: <https://dorl.net/dor/20.1001.1.23224347.1402.11.4.2.9>



* این مقاله یک مقاله با دسترسی آزاد است که تحت شرایط و ضوابط مجوز Creative Commons Attribution (CC BY) توزیع شده است.

نویسندگان ©

ناشر: دانشگاه جامع امام حسین (ع)

چکیده

باتوجه به اینکه بیشتر مسائل دنیای واقعی از قبیل تشخیص تقلب، شناسایی خطا، تشخیص ناهنجاری، تشخیص پزشکی و تشخیص بدافزار نامتوازن هستند، دسته‌بندی داده‌ها در مسائل نامتوازن به عنوان یکی از چالش‌های اصلی در حوزه داده‌کاوی، مورد توجه بسیاری از محققان و پژوهش‌گران قرار گرفته است. در یادگیری نامتوازن، معمولاً تعداد نمونه‌های یکی از دسته‌ها خیلی بیشتر از نمونه‌های دسته دیگر است و یا هزینه دسته‌بندی اشتباه در دودسته متفاوت است. شبکه‌های عصبی کانولووشنال به‌رغم موفقیت‌های چشمگیری که در دسته‌بندی داده‌ها دارند، در مسائل نامتوازن با مشکل مواجه می‌شوند چرا که آنها به‌صورت پیش‌فرض، توزیع دسته‌ها را متوازن و هزینه دسته‌بندی را مساوی در نظر گرفته می‌گیرند، از این‌رو در دسته‌بندی نامتوازن، نمی‌توان به نتایج قابل‌قبولی دست‌یافت؛ زیرا شبکه به سمت نمونه‌های آموزشی دسته بزرگ‌تر متمایل می‌شود که این موضوع سبب افزایش تعداد خطاها در تشخیص نمونه‌های مثبت می‌شود. یکی از راهکارهای کم‌هزینه برای غلبه بر نامتوازی داده‌ها در شبکه‌های عصبی کانولووشنال استفاده از تابع ضرر به نفع دسته اقلیت است، در این مقاله تابع ضرر جدیدی معرفی شده است که به‌صورت تدریجی و با پیشرفت آموزش، اهمیت دسته اقلیت را افزایش می‌دهد تا در انتهای آموزش به مقدار مشخص شده برسد و از اهمیت داده‌های دسته اکثریت بکاهد، این امر باعث می‌شود تا هم بتوانیم از قدرت آموزشی همه داده‌ها استفاده کنیم و هم از غلبه داده‌های دسته اکثریت جلوگیری کنیم. نتایج آزمایش روی سه مجموعه‌داده مصنوعی، تشخیص فعالیت‌های انسان و Cifar-10، همگرایی و کارایی روش پیشنهادی را نشان می‌دهند، روش پیشنهادی با روش‌های آدابوست مبتنی بر درخت تصمیم، شبکه کانولووشنال مبتنی بر آنتروپی متقابل و آنتروپی متقابل وزن‌دار، روش SMOTE و روش CNN تجمعی مقایسه شده است. به ترتیب با کسب دقت ۹۴/۶، ۹۲/۹۲ و ۶۹/۲۳ در سه مجموعه‌داده (Cifar-10 با نرخ نامتوازی ۵ درصد) توانست از دیگر روش‌ها پیشی بگیرد و دقت در مجموعه‌داده مصنوعی نسبت به روش سنتی آدابوست مبتنی بر درخت تصمیم، ۱۷/۷۲ بالاتر است.

کلید واژه‌ها: نامتوازی دسته‌ای، یادگیری عمیق، تابع ضرر، آنتروپی متقابل.

۱. مقدمه

نمونه کم، دسته اقلیت^۷ و به دسته‌های با نمونه زیاد، دسته اکثریت^۸ گفته می‌شود. راه‌حل‌ها برای برخورد با مساله نامتوازن دسته‌ای در شبکه‌های عمیق، به سه دسته تقسیم می‌شوند: ۱- راه حل‌های "مبتنی بر داده" [۲،۳] که تلاش دارند با کم یا زیاد کردن نمونه‌های برخی از دسته‌ها (افزایش نمونه‌های دسته اقلیت و یا کم کردن نمونه دسته اکثریت)، بر نامتوازن دسته‌ای غلبه کنند، ۲- راه حل‌هایی که با تغییر در ساختار مدل، روند آموزش شبکه، تابع ضرر و یا تغییر دادن هزینه‌ها در یادگیری نامتوازن به دنبال رسیدن به نتایجی به نفع دسته اقلیت هستند به راه حل‌های "مبتنی بر الگوریتم" [3،2] شهرت دارند، ۳- روش‌های ترکیبی که از هر دو روش مبتنی بر داده و مبتنی بر الگوریتم برای مقابله با نامتوازن داده استفاده می‌کنند [۴-۶]. در **Error!** *Reference source not found.* خلاصه‌ای از نحوه عملکرد برخی از روش‌های پیشنهادی با رویکردهای مبتنی بر داده، مبتنی بر الگوریتم و ترکیبی آورده شده‌است. در مواردی که داده‌های آموزشی زیاد باشد و یا روند آموزش زمان‌گیر باشد استفاده از روش‌های مبتنی بر الگوریتم اولویت دارد، یک روش الگوریتمی ساده و کم‌هزینه که در بیشتر موارد در یادگیری نامتوازن استفاده می‌شود، استفاده از تابع ضرر وزن دار است، که در مرحله‌ی آموزش شبکه به جریمه‌های مربوط به هر دسته، وزن ثابتی بین صفر تا یک اختصاص می‌دهد که این وزن در ارتباط معکوس با تعداد نمونه‌های هر دسته است و این وزن از ابتدا تا انتهای آموزش ثابت است، گرچه این تابع ضرر به خوبی به دسته اقلیت اهمیت می‌دهد اما از طرفی موجب می‌شود برخی از دسته‌ها که تعداد نمونه بالایی دارند و می‌توانند در یادگیری ویژگی‌های کلی توسط شبکه نقش بسزایی داشته باشند، وزن بسیار پایینی پیدا کنند و کارایی شبکه کاهش پیدا کند. در این مقاله برای غلبه بر نامتوازن داده‌ها یک روش مبتنی بر الگوریتم ارائه شده است، که سعی در رفع مشکلات روش آنتروپی متقابل وزن دار^۹ WCE در برخورد با داده‌های نامتوازن دارد. ما با استفاده از تابع ضرر WCE و CE روی داده‌های نامتوازن، با نرخ نامتوانی مختلف، و با اندازه‌گیری معیار FI-score روی داده‌های دسته اقلیت در شکل دریافتیم زمانی که نرخ نامتوانی دسته‌ای بالاست، WCE نه تنها نتایج بهتری نسبت به CE نمی‌دهد بلکه در برخی موارد نتایج حاصل از CE نسبت به WCE بهتر است (مگر زمانی که داده‌های دسته اقلیت به قدری پایین باشد که با استفاده از تابع آنتروپی متقابل استاندارد به عنوان نویز شناخته شود، در اینجا $p=2$) همانطور که مشاهده می‌شود در نرخ نامتوانی ۱۰، ۵، و ۳ درصد CE کارایی بهتری از WCE از خود نشان داده است. ما در این مقاله تابع ضرر وزن داری به نام IWCE

دسته‌بندی یا پیش‌بینی نامتوازن ممکن است هرجایی اتفاق بیفتد، چه در کاربردهای زندگی واقعی و چه در تحلیل داده‌های علمی، یادگیری نامتوازن^۱ به این معناست که یکی از دسته‌ها، احتمال پیشین متفاوت دارد (داده‌های نامتوازن^۲) و یا هزینه دسته‌بندی اشتباه در دودسته متفاوت است (یادگیری حساس به هزینه^۳) و یا هر دو. اغلب مسائل دنیای واقعی غیرمتوازن هستند، یک پزشک وقتی از بیمار برای تشخیص یک بیماری خطرناک آزمایش می‌گیرد با یک تصمیم حساس به هزینه روبرو است، یک نتیجه مثبت اشتباه^۴ حداکثر می‌تواند منجر به آزمایش‌های بیشتر شود؛ اما یک نتیجه منفی اشتباه^۵ می‌تواند مرگ‌بار باشد. یک دوربین نظارتی که باید درگیری خیابانی را تشخیص دهد، باید تصاویر دوربین نظارتی را بررسی کند تا ببیند درگیری خیابانی اتفاق افتاده یا خیر. این مسئله یک مسئله یادگیری نامتوازن است که بیشتر نمونه‌ها جزو دسته منفی هستند، چرا که تصاویری که در آنها درگیری اتفاق افتاده باشد بسیار کم است و همچنین برای مسئله تشخیص حوادث طبیعی در ویدئوها نیز مسئله به همین صورت است. این مسئله برای سیستم‌های بانکی تشخیص تقلب در کارت‌های اعتباری نیز وجود دارد، تراکنش‌های غیرقانونی نسبت به تراکنش‌های صحیح و قانونی بسیار کم است. همین‌طور هزینه پیش‌بینی منفی اشتباه بیشتر از هزینه پیش‌بینی مثبت اشتباه است، وقتی فعالیت‌های مشکوک در حساب‌های مشتریان را شناسایی می‌کنیم، اگر اشتباهی رخ دهد، ترجیح می‌دهیم مشتری کمی از تشخیص اشتباه ما ناراحت شود (مثبت اشتباه) تا اینکه مشتری قربانی کلاهبرداری (منفی اشتباه) شود. همچنین مسئله تشخیص شایعه در شبکه‌های اجتماعی یک مسئله نامتوازن دسته‌ای است؛ چراکه تعداد پیام‌های عادی بسیار بیشتر از پیام‌های شایعه است [1].

امروزه شبکه‌های عمیق باعث پیشرفت‌های خیره‌کننده‌ای در تکنولوژی‌های مورد استفاده در اطراف ما شده‌اند، مثل تشخیص پزشکی، ماشین‌های خودراننده، ترجمه‌های ماشینی، تشخیص صدا، تشخیص تصاویر و غیره. اما مانند دیگر تکنیک‌های یادگیری ماشین، هنگام کار با داده‌های نامتوازن تمایل به سمت دسته با داده‌های زیاد خواهند داشت. در این پژوهش به دنبال ارائه راهکاری مناسب و کم هزینه برای شبکه‌های یادگیری عمیق در مواجهه با مساله داده‌های نامتوازن یا همان نامتوانی دسته-ای^۶ هستیم، در این مسائل داده‌های نامتوازن، به دسته‌های با

¹ asymmetric learning

² Imbalanced data

³ Cost sensitive

⁴ false positive

⁵ false negative

⁶ Class Imbalanced Learning

⁷ Minority

⁸ Majority

⁹ weighted cross entropy

نمونه‌هایی هستند که به درستی در دسته مثبت دسته‌بندی شده‌اند و TN تعداد نمونه‌هایی هستند که به درستی در دسته منفی دسته‌بندی شده‌اند، FN و FP به ترتیب تعداد نمونه‌هایی است که به طور اشتباه در دسته منفی و مثبت دسته‌بندی شده‌اند. معیارهایی که در ادامه در این بخش (رابطه‌های ۱ تا ۷) به آن اشاره شده، از ماتریس درهم‌ریختگی به دست آمده است.

ماتریس درهم ریختگی

	واقعی Positive (1)	واقعی Negative (0)
دسته بندی شده Positive (1)	True Positives (TPs)	False Positives (FPs)
دسته بندی شده Negative (0)	False Negatives (FNs)	True Negatives (TNs)

جدول (۱). ماتریس درهم ریختگی

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$error\ rate = 1 - Accuracy$$

این معیارها یک روش ساده برای بیان کارایی دسته‌بند روی مجموعه داده‌ها را بیان می‌کند. به هر حال در شرایط خاص می‌تواند فریب‌دهنده باشد و نسبت به تغییر در داده‌ها بسیار حساس است. در ساده‌ترین حالت، اگر ترکیب مجموعه داده‌ها، ۵ درصد دسته اقلیت و ۹۵ درصد دسته اکثریت باشد، یک راهکار ساده‌ای که همه نمونه‌ها را در دسته اکثریت دسته‌بندی می‌کند، می‌تواند دقت ۹۵ درصد را به دست آورد. در ظاهر دقت ۹۵ درصد روی کل داده‌ها عالی به نظر می‌رسد، با این حال نمی‌تواند منعکس‌کننده این حقیقت باشد که ۰ درصد از نمونه‌های دسته اقلیت شناسایی شده‌اند. به این دلیل است که می‌گوییم، معیار دقت در این حالت اطلاعات مناسبی را درباره کارایی دسته‌بند با توجه به نوع دسته‌بندی ارائه نمی‌دهد. با داده‌های نامتوازن، هنگامی که معیارهای ارزیابی به توزیع داده‌ها حساس هستند، تجزیه و تحلیل نسبی مشکل می‌شود. به جای دقت، دیگر معیارهای ارزیابی که معمولاً اتخاذ می‌شوند عبارت است از: صحت^۲ (رابطه ۱)، بازخوانی^۳ (رابطه ۲)، $f\text{-measure}(F1\text{-score})$ (رابطه ۳):

$$precision = \frac{TP}{TP+FP} \quad (2)$$

$$recall = \frac{TP}{TP+FN} \quad (3)$$

$$f\text{-measure} = \frac{(1+\beta^2) \times precision \times recall}{\beta^2 \times precision + recall} \quad (4)$$

$(incremental\ weighted\ cross\ entropy)$ پیشنهاد کرده‌ایم که به بهبود یادگیری نامتوازن و افزایش معیارها مخصوصاً زمانی که نرخ نامتوازنی دسته‌ها زیاد است، کمک می‌کند. در این روش در ابتدا به همه دسته‌ها وزن یکسانی اختصاص داده می‌شود و سپس در هر اپوک وزن دسته اکثریت و اقلیت براساس تعداد نمونه‌های آن به صورت تدریجی و براساس α که هاپیر پارامتر قابل تنظیم است تغییر می‌کند تا اینکه در انتهای آموزش وزن هر دسته به مقدار از پیش محاسبه شده می‌رسد که همان وزن دسته‌ها در روش انتروپی متقابل وزن دار (WCE (weighted cross entropy است. α سرعت تغییر وزن را مشخص می‌کند.

همان‌طور که از شکل ۱ مشخص است، $IWCE$ کارایی بهتر از CE و WCE از خود نشان داده است. ما روش پیشنهادی خود را در بخش ۵ به طور مفصل معرفی کرده و در بخش ۶ روی سه مجموعه داده داده‌های مصنوعی، ردیابی با موبایل و $cifar-10$ اجرا و کارایی آن را بررسی کرده‌ایم، همچنین در این بخش کارایی روش پیشنهادی ما در مقایسه با دیگر تکنیک‌های مقابله با نامتوازنی داده‌ای سنجیده می‌شود. نوآوری‌های روش پیشنهادی در مواجهه با نامتوازنی داده‌ها:

- ۱- استفاده از شبکه‌های کانولوشنال برای استخراج ویژگی‌های سطح بالا از ویژگی‌های اولیه به طوری که برای مجموعه داده‌های نامتوازن بیشترین جداکنندگی را داشته باشد.
- ۲- رفع مشکل از دست رفتن اطلاعات مفید داده‌های دسته اکثریت ضمن افزایش تاثیر داده‌های کلاس اقلیت با جلوگیری از کاهش وزن شدید آن بوسیله معرفی فرایند وزن دهی مرحله‌ای در رابطه (۱۴) و (۱۵).
- ۳- معرفی α به عنوان یک هاپیر پارامتر برای کنترل شتاب تغییر وزن با افزایش دوره‌های آموزش شبکه (با افزایش اپوک)

در ادامه در بخش ۲ در ابتدا معیارهای ارزیابی داده‌های نامتوازن بررسی می‌شوند سپس در بخش ۳ و ۴ مرور کلی بر روش‌های مقابله با نامتوازنی دسته‌ای در شبکه‌های عمیق خواهیم داشت، در بخش ۴ به بررسی کارهای انجام شده در زمینه تغییر تابع ضرر می‌پردازیم، در بخش ۵ راهکار پیشنهادی معرفی می‌شود. نتایج و مقایسه‌ها در بخش ۶ گنجانده شده و بخش ۷ نتیجه گیری و پیشنهادها آتی قرار داده شده است.

۲. معیارهای ارزیابی برای داده‌های نامتوازن

ماتریس درهم‌ریختگی^۱ که در جدول (۱) نشان داده شده است، نتایج یک دسته‌بندی دودویی را نمایش می‌دهد. TP تعداد

^۲ precision
^۳ recall

^۱ confusion matrix

۳. یادگیری عمیق برای داده‌های نامتوازن

امروزه یادگیری عمیق به دلیل اعمال ایده‌های جدید، الگوریتم‌های جدید و طراحی معماری‌های شبکه‌ای جدیدی پیشرفت‌های چشمگیری در همه زمینه‌ها به‌خصوص در زمینه تشخیص تصویر و شناسایی اشیا داشته باشد، با وجود همه این پیشرفت‌ها مشکلاتی را از اجداد خود یعنی شبکه‌های عصبی سنتی به ارث برده‌اند که از آن میان می‌توان به مشکلات شبکه‌های عصبی در برخورد با داده‌های نامتوازن اشاره کرد. آناند و همکاران [۷] که تأثیرات نامتوازنی دسته‌ای در الگوریتم انتشار روبه‌عقب^۱ در شبکه‌های عصبی را بررسی کردند، نشان دادند که در سناریوهای دسته نامتوازن، طول مؤلفه گرادیان دسته اقلیت بسیار کمتر از مؤلفه گرادیان دسته اکثریت است. به‌عبارت‌دیگر، دسته اکثریت اساساً بر گرادینانی که وزن مدل را به‌روزرسانی می‌کند، تسلط دارد. این موضوع در تکرارهای اولیه خیلی سریع خطای دسته اکثریت را کاهش می‌دهد، اما اغلب خطای دسته اقلیت را افزایش می‌دهد و باعث می‌شود شبکه در یک حالت همگرایی کندگیر کند.

در سال‌های اخیر حل مشکل داده‌های نامتوازن در شبکه‌های عمیق بسیار مورد توجه و علاقه محققان قرار گرفته است. کارهای انجام شده در زمینه یادگیری نامتوازن در شبکه‌های عصبی عمیق، بر اساس تکنیک مورد استفاده برای حل یادگیری نامتوازن به سه دسته روش‌های مبتنی بر داده، روش‌های مبتنی بر الگوریتم و روش‌های ترکیبی تقسیم می‌شوند که در بخش ۱ نحوه عملکرد این سه روش آورده شده است.

از بین روش‌های مبتنی بر داده می‌توان به روش‌هایی مانند^۲ $[ROS, RUS]$ ، روش نمونه‌گیری پویا [۱۰] که نرخ نمونه‌گیری را با توجه به عملکرد هر دسته تنظیم می‌کند و یادگیری دو فاز [۱۱، ۱۲]، اشاره کرد. برخلاف روش‌های داده‌ای، روش‌های الگوریتمی، داده‌های آموزشی را تغییر نمی‌دهند و به مراحل پیش‌پردازش نیاز ندارند. در مقایسه با ROS ، که پرتفردارترین روش مبتنی بر داده است، روش‌های الگوریتمی کمتر بر زمان آموزش تأثیر می‌گذارند. به نظر می‌رسد که روش‌های الگوریتمی می‌توانند برای مسائل داده‌های عظیم بهتر باشند. به استثنای تعریف هزینه‌های دسته‌بندی اشتباه، روش‌های الگوریتمی به تنظیماتی کمتری نیاز دارند. در ادامه به بررسی چند روش الگوریتمی می‌پردازیم.

که β ضریبی است که برای تنظیم اهمیت دقت نسبت به بازخوانی (معمولاً یک است) بکار برده می‌شود. صحت، میزان دقیق بودن است (برای مثال نمونه‌هایی که برچسب مثبت خورده‌اند تا چه حد درست برچسب‌گذاری شده‌اند) درحالی‌که بازخوانی معیار کامل بودن است (برای مثال، چه تعداد از نمونه‌های دسته مثبت درست برچسب‌گذاری شده‌اند). برعکس دقت و نرخ خطا که هر دو به توزیع مجموعه داده حساس بودند، صحت و بازخوانی هر دو به توزیع حساس نیستند. یک بررسی ساده‌روی رابطه‌های ۲ و ۳ نشان می‌دهد صحت به توزیع دسته حساس است اما بازخوانی نه. با این حال در صورتی که به‌درستی استفاده شود، صحت و بازخوانی می‌تواند به‌صورت کارا، کارایی دسته‌بند را در روند یادگیری نامتوازن ارزیابی کند. به‌خصوص معیار F - $Measure$ (که با F -score شناخته می‌شود) با در نظر گرفتن نسبت اهمیت بازخوانی به صحت که توسط کاربر مشخص می‌شود، صحت و بازخوانی را ترکیب می‌کند. F - $Measure$ درک عمیق‌تری از کارایی دسته‌بند نسبت به معیار صحت ارائه می‌کند، در مواردی که با به‌جای مسئله دو دسته‌ای با مسئله چند دسته‌ای روبرو هستیم، اگر ماتریس درهم‌ریختگی را CM بنامیم برای دسته i $recall_i$ (بازخوانی دسته i ام) به‌صورت تقسیم عنصر $CM_{i,i}$ بر مجموع عناصر ستون i ام تعریف می‌شود (رابطه ۶) و $precision_i$ (صحت دسته i ام) به‌صورت تقسیم عنصر $CM_{i,i}$ بر مجموع عناصر ستون i ام تعریف می‌شود و معیار F - $Measure$ برای تک‌تک دسته‌ها به‌صورت جداگانه محاسبه می‌شود. معیار دیگر $Balanced accuracy$ یا دقت متوازن شده است که به‌طور گسترده از آن در یادگیری نامتوازن برای بررسی کارایی دسته‌بندها استفاده می‌شود، این رابطه در واقع میانگین دقت هر یک از دسته‌ها و از آنجاکه مجموع ستون i ام ماتریس درهم‌ریختگی برابر با تعداد نمونه‌های دسته i ام است می‌توان دقت متوازن شده را به‌صورت میانگین $recall$ دسته‌ها بیان کرد، در این مقاله برای مقایسه کارایی روش پیشنهادی از معیارهای صحت، بازخوانی و $F1$ -score (به این معنی است که β یک در نظر گرفته شده است) و $Balanced accuracy$ استفاده شده است. در رابطه (۵) و (۶) CM نشان‌دهنده ماتریس درهم‌ریختگی است و در رابطه (۷) C برابر تعداد دسته‌ها و acc_i برابر دقت دسته‌بند در هر دسته است.

$$precision_i = \frac{CM_{i,i}}{\sum_j CM_{i,j}} \quad (5)$$

$$recall_i = \frac{CM_{i,i}}{\sum_j CM_{j,i}} \quad (6)$$

$$balanced accuracy = \frac{\sum_i acc_i}{C}, acc_i = recall_i \quad (7)$$

¹ backpropagation

² Random over sampling

³ Random under sampling

۴. کارهای انجام شده

در این بخش به بررسی پژوهش‌هایی می‌پردازیم که با تغییر در الگوریتم شبکه و یا تغییر تابع ضرر به نفع کلاس سعی در کاهش اثرات نامطلوب نامتوازنی دسته‌ای داشته‌اند. اقلیت وانگ و همکاران [۱۳] در ابتدا نشان دادند که تابع ضرر میانگین مربع خطا (MSE)، در مواردی که نامتوازنی دسته‌ای بالاست، خطای دسته اقلیت را به سختی منعکس می‌کند، زیرا زیاد بودن نمونه‌های منفی باعث می‌شود آنها بر تابع ضرر غالب باشند. آنها سپس دو تابع ضرر جدید پیشنهاد کردند که نسبت به خطاهای دسته اقلیت حساس‌تر هستند، میانگین خطای اشتباه (MFE) و میانگین مربع خطای اشتباه ($MSFE$). توابع ضرر پیشنهادی با تقسیم MSE به دو جزء، میانگین خطای مثبت اشتباه (FPE) و میانگین خطای منفی اشتباه (FNE) بدست آمده است

برای غلبه بر عدم توازن شدید بین دسته پیش‌زمینه و پس‌زمینه (اشیا) در شناسایی اشیا، لین و همکاران [۱۴] تابع ضرر کانونی^۱ را معرفی کردند، آنها تابع ضرر آنتروپی متقابل (CE) را دوباره بازنویسی کردند به طوری که تأثیر نمونه‌هایی که به راحتی دسته‌بندی شده‌اند، روی تابع ضرر را کاهش دهند. این ایده با ضرب تابع ضرر CE در یک عامل تعدیل کننده، $\gamma^{\alpha_i (1-p)}$ حاصل می‌شود. هایپر پارامتر $\gamma \geq 0$ میزان کم وزنی مثال‌های آسان را تنظیم می‌کند و $\alpha_i \geq 0$ وزنی دسته‌ای است که برای افزایش اهمیت دسته اقلیت استفاده می‌شود. احمد و عبدالفتاح در [۲۱] این روش را در شبکه‌های ResNet و DensNet برای تشخیص کوید ۱۹ از روی تصاویر اشعه x بکار بردند.

برخی از پژوهش‌ها از روش‌های حساس به هزینه استفاده کردند به طوری که به خطاهای منفی اشتباه^۲ هزینه بیشتر از هزینه خطاهای مثبت اشتباه^۳ داده می‌شود [۱۵، ۱۶]. در [۲۲] طاهرخانی و همکاران، روش آدابوست مبتنی بر CNN ها برای داده‌های نامتوازن ارائه دادند که چندین CNN را به‌عنوان دسته‌بند پایه با روش SAMME [۲۳] با هم ترکیب می‌کند و هر CNN جدید با استفاده از یادگیری انتقالی

وزن‌های خود را از CNN قبلی کپی می‌کند و سپس با داده جدید تنظیم^۴ می‌شود و سپس نتایج ترکیب می‌شوند. اولام و همکاران روشی را در [۲۴] برای پیش‌بینی کووید ۱۹ ارائه داده است که از یادگیری انتقالی در کنار روش افزایش داده SMOTE استفاده کرده است.

در [۲۵]، خالویی و همکاران برای تشخیص رویداد های نادر در ویدیو در مجموعه داده UCSDped1 و UCSDped2 شبکه‌ای مشابه شبکه مولد تخصصی^۵ طراحی کردند و با کمک تکنیک نرمالسازی دسته‌ای، رویدادهای غیر رایج را از رویداد رایج مجزا می‌کردند. بودا و همکاران [۱۲] برای بهبود عملکرد کلی، آستانه خروجی CNN را تنظیم کردند. نویسندگان برای استفاده از حد آستانه، خروجی‌های شبکه برای هر دسته را بر احتمال پیشین تخمینی دسته تقسیم کردند، و به طور موثر احتمال دسته‌بندی نادرست نمونه‌های گروه اقلیت را کاهش دادند.

دینگ و همکاران [۱۹] برای تعیین اینکه بررسی کنند آیا شبکه‌های عمیق‌تر در داده‌های نامتوازن عملکرد بهتری دارند، با معماری‌های بسیار عمیق CNN، به عنوان مثال، ۵۰ لایه، آزمایش انجام دادند. نویسندگان مشاهده کردند که سطوح خطای شبکه‌های عمیق‌تر نسبت به شبکه‌های با اندازه کوچکتر از کیفیت بهتری برای همگرایی آموزش برخوردار هستند [۲۶، ۲۷]. آنها ادعا می‌کنند که شبکه‌های بزرگتر دارای مینیمم محلی با کارایی مناسب هستند و با کاهش گرادیان می‌توان رسیدن به راه‌حل‌های قابل قبول را آسان تر کرد.

Error! Reference source not found. در

خلاصه ای از روش های ذکر شده برای مقابله با نامتوازنی دسته ای با نوع شبکه استفاده شده و نوع روش بکاررفته، آورده شده است.

^۴ fine tune

^۵ generative adversarial network(GAN)

^۱ Focal loss

^۲ false negative errors

^۳ false positive errors

جدول (۲). خلاصه ای از روش های استفاده شده در یادگیری عمیق برای غلبه بر نامتوازن دسته ای [۲]

شرح	نوع روش	نوع شبکه	روش
<i>ROS</i> را روی داده های دسته اقلیت اعمال می کند تا به توازن برسد.	داده ای	<i>CNN</i>	<i>ROS</i> [۱۲،۸]
<i>RUS</i> را روی داده های دسته اکثریت اعمال می کند تا به توازن برسد.	داده ای	<i>CNN</i>	<i>RUS</i> [۱۲]
پیش آموزش با <i>RUS</i> یا <i>ROS</i> انجام می شود، سپس تنظیم وزن ها با همه داده ها انجام می شود	داده ای	<i>CNN</i>	<i>Two-phase learning</i> [۱۲،۹]
وزن نمونه ها حین آموزش و بر اساس <i>f1-score</i> دور قبلی تنظیم می شود	داده ای	<i>CNN</i>	<i>Dynamic sampling</i> [۱۰]
تابع ضرر جدید به دسته های مثبت و منفی اجازه می دهد به نسبت مساوی در تابع ضرر همکاری داشته باشند.	الگوریتم	<i>MLP</i>	<i>MFE and MSFE loss</i> [۱۳]
تابع ضرر جدید وزن نمونه هایی که به راحتی دسته بندی می شوند را کم می کند و تأثیر آنها در تابع ضرر نهایی را کاهش می دهد.	الگوریتم	<i>CNN</i>	<i>Focal loss</i> [۱۴]
تابع ضرر اصلاح شده است تا یک ماتریس هزینه از پیش تعریف شده را در آن بگنجانند.	الگوریتم	<i>MLP</i>	<i>CSDNN</i> [۱۵]
ماتریس هزینه از طریق پس انتشار یاد گرفته می شود و در لایه خروجی گنجانده می شود.	الگوریتم	<i>CNN</i>	<i>CoSen CNN</i> [۱۶]
ماتریس هزینه از طریق الگوریتم تکاملی یاد گرفته می شود و در لایه خروجی قرار می گیرد.	الگوریتم	<i>DBN</i>	<i>CSDBN-DE</i> [۱۷]
آستانه تصمیم گیری با تقسیم احتمالات خروجی بر احتمال پیشین دسته تنظیم می شود	الگوریتم	<i>CNN</i>	<i>Threshold moving</i> [۱۲]
مرکز دسته ها در فضای ویژگی عمیق محاسبه می شود و به روش <i>KNN</i> دسته ها از هم متمایز می شوند.	الگوریتم	<i>CNN</i>	<i>Category centers</i> [۱۸]
شبکه های <i>CNN</i> با عمق بالای ۵۰ برای امتحان کردن سرعت همگرایی بکار برده می شود.	الگوریتم	<i>CNN</i>	<i>Very-deep NNs</i> [۱۹]
تابع ضرر سه تایی و نمونه برداری پنج تایی ویژگی های متمایزکننده بهتری تولید می کنند.	ترکیبی	<i>CNN</i>	<i>LMLE</i> [۴]
با استفاده از <i>KNN</i> و ضرر <i>micro-cluster</i> نمونه های دسته اقلیت در فضای ویژگی عمیق افزایش پیدا می کند	ترکیبی	<i>CNN</i>	<i>DOS</i> [۲۰]
تابع ضرر <i>CRL</i> و کاوش نمونه های سخت، ویژگی های متمایزکننده تری تولید می کند.	ترکیبی	<i>CNN</i>	<i>CRL loss</i> [۶]

$$D_{KL}(p \parallel q) = \sum_{i=1}^n q(x_i) \log_e \left(\frac{q(x_i)}{p(x_i)} \right) \quad (۸)$$

بنابراین، واگرایی KL را می توان به صورت رابطه (۹) ساده کرد:

$$\begin{aligned} D_{KL}(p \parallel q) &= \sum_{i=1}^n q(x_i) \log_e \left(\frac{q(x_i)}{p(x_i)} \right) \\ &= \sum_{i=1}^n q(x_i) \log_e(q(x_i)) - \sum_{i=1}^n q(x_i) \log_e(p(x_i)) \end{aligned}$$

۵. راهکار پیشنهادی

تابع ضرر در واقع روشی برای بیان میزان تفاوت پیش بینی شبکه برای نمونه x_i ($p(x_i)$) و مقدار واقعی هدف است ($q(x_i)$) است، تابع KL یک تابع واگرایی است که تفاوت دو تابع احتمال $p(x)$ و $q(x)$ را محاسبه می کند، واگرایی KL را می توان به صورت رابطه (۸) نشان داد.

تعداد نمونه‌های دسته اقلیت نسبت به اکثریت بسیار پایین باشد که این باعث می‌شود طبق رابطه ۸، وزن دسته اکثریت بسیار کاهش پیدا کند و مقدار ناچیزی شود و در واقع تأثیر داده‌های اکثریت در آموزش بسیار کم شود. فرض کنید تعداد نمونه‌های دسته اقلیت ۱۰۰ و تعداد نمونه‌های دسته اکثریت ۲۰۰۰ باشد وزن دسته اقلیت برابر ۱۰.۵ و وزن دسته اکثریت ۰.۵۲۵ خواهد بود و بعد از نرمال‌سازی ۱ و ۰.۰۵ خواهد بود و در واقع دسته‌بند ما با داده‌های محدود دسته اقلیت آموزش می‌بیند و وزن اطلاعات داده‌های مفید اکثریت بسیار ناچیز خواهد بود و این باعث کاهش کارایی کلی دسته‌بند خواهد شد

در واقع از آنجاکه توزیع دسته اکثریت و اقلیت از یکجا می‌آید [۳۰] زمانی که نمونه‌های دسته اقلیت بسیار پایین است، نمونه‌های دسته اکثریت به یادگیری شبکه برای نمونه‌های دسته اقلیت کمک می‌کند و کاهش بیش از اندازه وزن آن باعث می‌شود نمونه‌های اقلیت نیز به خوبی یاد گرفته نشود و همانطور که از جدول ۳ برای نرخ‌های نامتوازنی ۲ درصد، ۵ درصد و ۱۰ درصد مشخص است به خاطر کاهش قدرت داده‌های آموزشی، مدل دچار بیش‌برازش می‌شود که بیش‌برازش با افزایش نامتوازنی افزایش پیدا می‌کند. یک راهکار برای جلوگیری از این موضوع می‌تواند استفاده از لگاریتم وزن‌ها به جای خود وزن‌ها باشد. اما حتی برای دو عدد بسیار متفاوت، لگاریتم اعدادی نزدیک به هم ارائه می‌دهد و نمی‌تواند نشان دهنده میزان اهمیت یک دسته از نمونه‌ها باشد. برای اینکه هم بتوان از قدرت آموزشی داده‌های دسته اکثریت در حین آموزش استفاده کرد و هم از بیش‌برازش مدل روی داده‌های آموزشی جلوگیری کرد، در این مقاله تابع ضرر وزن دار افزایشی پیشنهاد شده است.

$$\begin{aligned} &= -H(q) - \sum_{i=1}^n q(x_i) \log_e(p(x_i)) \\ &= H(q.p) - H(q) \end{aligned} \quad (9)$$

که $H(q)$ به صورت رابطه (۱۰) تعریف می‌شود:

$$H(p) = - \sum_{i=1}^n q(x_i) \log_e(q(x_i)) \quad (10)$$

درواقع $H(q.p)$ آنتروپی متقابل p و q است. از آنجایی که $H(q(x))$ ثابت است و تغییری نمی‌کند، مینیمم کردن آنتروپی متقابل باعث مینیمم شدن واگرایی KL بین پیش‌بینی شبکه و مقدار واقعی هدف می‌شود و شبکه به سمت جواب درست پیش خواهد رفت.

$$\begin{aligned} \text{Cross entropy}(p,q) &= \\ H(p,q) &= - \sum_{i=1}^n p(x_i) \log_e(q(x_i)) \end{aligned} \quad (11)$$

در بخش ۱-۴ آنتروپی متقابل وزن دار را بررسی می‌کنیم.

۱-۵. تابع آنتروپی متقابل وزن دار

یک راه‌حل رایج و کم‌هزینه برای افزایش تأثیر داده‌های دسته اقلیت در مرحله آموزش و تغییر تابع ضرر به نفع دسته اقلیت، استفاده از آنتروپی متقابل وزن دار است که در رابطه (۱۲) نشان داده شده است و وزن هر دسته در این رابطه، ارتباطی معکوس با تعداد نمونه‌های هر دسته دارد [۲۸،۲۹].

$$\begin{aligned} \text{weighted cross entropy} &= \\ &= - \sum_{i=1}^n w_i t_i \log(p_i) \end{aligned} \quad (12)$$

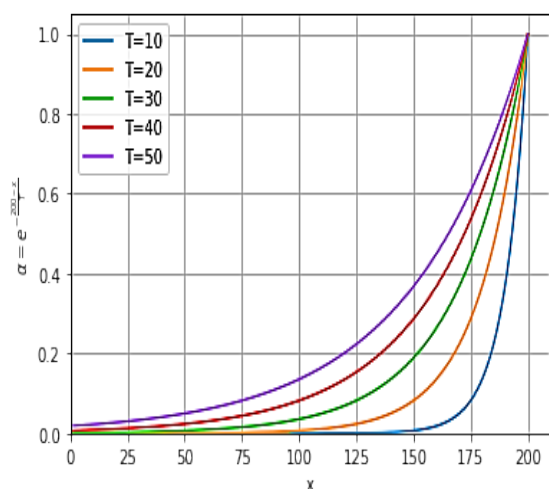
w_i برای دسته i ام در $\text{Weighted Cross Entropy}$ (به اختصار WCE) به صورت زیر (رابطه ۱۳) محاسبه می‌شود:

$$w_i = \frac{\sum_{i=1}^c n_i}{c \times n_i} \quad (13)$$

c تعداد دسته‌ها و n_i تعداد نمونه‌ها در دسته i است. این وزن‌ها قبل از استفاده در تابع آنتروپی متقابل وزن دار نرمال‌سازی می‌شوند. ما برای بررسی رفتار WCE از مجموعه داده سیفار استفاده کردیم و نمونه‌های نیمی از دسته‌ها را با نرخ‌های نامتوازنی مختلف، کاهش دادیم و سپس معیار minority average F1-score را در نرخ‌های نامتوازنی مختلف اندازه گرفتیم و با تابع ضرر CE مقایسه کردیم. نتایج همان‌طور که از نتایج در شکل ۱ مشخص است، با اینکه در نرخ‌های نامتوازنی متوسط (۶۰، ۴۰ و ۲۰ درصد) WCE بهتر از CE عمل می‌کند؛ اما زمانی که نامتوازنی دسته‌ای افزایش پیدا می‌کند و زیاد می‌شود، می‌بینیم که CE عملکردی بهتر از WCE از خود نشان می‌دهد، (به جز در نامتوازنی ۲ درصد که CE داده‌های دسته اقلیت را به‌عنوان نویز می‌شناسد). اما علت چیست؟ مشکل زمانی است که

شکل (۱) بررسی معیار $F1\text{-score}$ برای سه تابع ضرر CE .

$IWCE$ و WCE



شکل (۱). رسم α برای T های مختلف

که در این رابطه x شماره ایپوک و T ضریب ثابتی است که در ابتدای آموزش به عنوان هایپر پارامتر مشخص می‌شود، در شکل ۳، α برای T های مختلف رسم شده است. طبق آزمایش‌های ما و با تکنیک T grid search مقادیر T در حدود ۳۰ الی ۵۰ مناسب است، بر طبق آزمایشات انجام شده، در T های پایین تر به دلیل تغییرات ناگهانی همگرایی شبکه با مشکل مواجه می‌شود. در شکل ۴، منحنی دو وزن فرضی برای دسته اقلیت ($w=0.9$) و اکثریت ($w=0.09$) نشان داده شده است. در تکرار های اولیه وزن دو دسته برابر است و با هر تکرار کم کم وزن دو دسته از هم فاصله می‌گیرد تا اینکه در تکرار نهایی وزن‌ها برابر با مقدار w_i (رابطه ۱۳) می‌شود. تغییرات مرحله‌ای وزن‌ها می‌تواند در هر ایپوک یا در هر n ایپوک یکبار انجام شود الگوریتم تابع ضرر افزایشی در جدول ۴ نشان داده شده است. تابع ضرر وزن دار افزایشی یا IWCE را می‌توانیم به صورت رابطه (۱۶) تعریف کنیم:

$$\text{incremental weighted cross entropy} = -\sum_1^n (\alpha w_i + (1 - \alpha) w_i^0) p_i \log(p_i) \quad (16)$$

که w_i از رابطه (۱۳) به دست می‌آید و w_i^0 وزن انتخابی در اولین ایپوک است.

۵-۲ تابع ضرر وزن دار افزایشی^۱

ما در این مقاله تابع ضرر وزن دار افزایشی را معرفی می‌کنیم. این تابع ضرر در طول آموزش تغییر می‌کند به این صورت که در ابتدای آموزش و در ایپوک‌های ابتدایی به همه دسته‌ها، وزن یکسانی می‌دهد و بدین ترتیب دسته‌بند ما به خوبی روی داده‌های آموزشی دسته اکثریت آموزش می‌بیند و با افزایش ایپوک‌ها به صورت تدریجی وزن دسته اکثریت کاهش و وزن دسته اقلیت افزایش می‌یابد تا اینکه در نهایت وزن دسته‌ها به w_i در رابطه ۱۳ می‌رسند. وزن جدید تابعی از α است که α خود براساس شماره ایپوک تغییر می‌کند. این وزن جدید در رابطه (۱۴) نشان داده شده است.

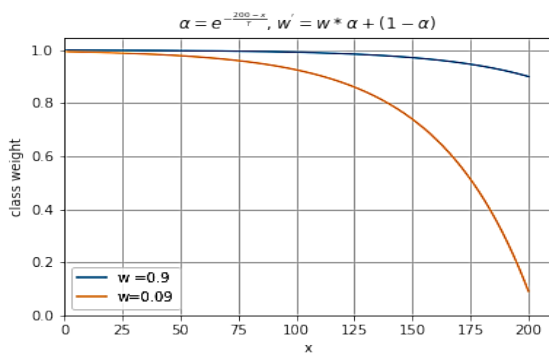
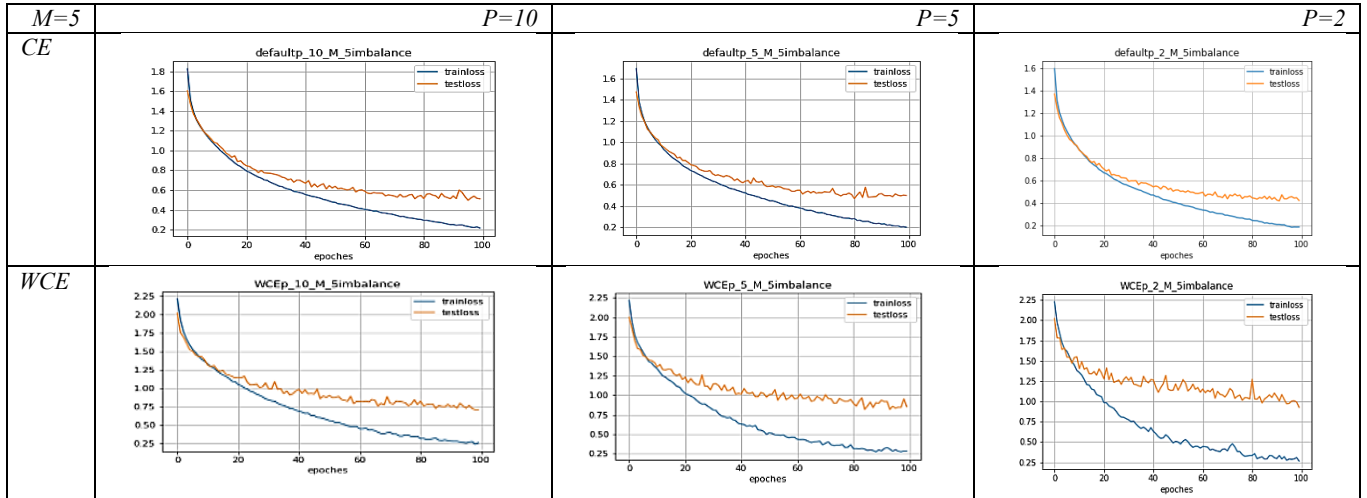
$$w_i' = \alpha w_i + (1 - \alpha) w_i^0 \quad (14)$$

که در این رابطه w_i^0 وزن اولیه دسته‌ها و w_i وزنی است که در رابطه ۱۳ مشخص شده و در ایپوک های نهایی می‌خواهیم به این وزن برسیم. در اینجا ما w_i^0 را که در واقع وزن دسته‌ها در ایپوک اولیه است برای همه دسته‌ها مساوی و برابر ۱ در نظر گرفته‌ایم. مقدار α از صفر تا یک تغییر می‌کند به این صورت که در ابتدای آموزش ۰ و در انتهای آموزش ۱ است. با پیشرفت آموزش به صورت مرحله‌ای با افزایش α ، وزن‌ها از مقدار مساوی فاصله می‌گیرند و به w_i نزدیک‌تر می‌شود و به تدریج وزن دسته اقلیت افزایش و وزن دسته اکثریت کاهش می‌یابد. پارامتر α ، که از رابطه (۱۵) پیروی می‌کند، پارامتر تنظیم کننده سرعت تغییر w_i است. این پارامتر از دو متغیر تاثیر می‌پذیرد، متغیر x و متغیر T ، متغیر x در واقع شماره ایپوک آموزش است که با پیشرفت آموزش افزایش پیدا می‌کند و باعث می‌شود وزن از یک ایپوک به ایپوک بعدی به آرامی تغییر کند و متغیر T که در واقع وظیفه تنظیم سرعت برعهده این متغیر است، یک کوچک تمام تغییرات را به ایپوک های آخر محول می‌کند و T بزرگ که از ابتدای آموزش تغییرات را شروع می‌کند.

$$\alpha = e^{-\frac{(200-x)}{T}} \quad (15)$$

¹ Incremental Weighted Cross Entropy

جدول (۳). بررسی نمودار تابع ضرر CE و WCE



شکل (۳). منحنی کاهش برای دو وزن فرضی دسته اقلیت و اکثریت در ۲۰۰ تکرار

۶. پیاده‌سازی و بررسی نتایج

برای بررسی تابع ضرر پیشنهادی، آن را روی سه مجموعه داده مختلف بررسی کردیم و نتایج را با روش‌های مشهور مقابله با نامتوازنی دسته‌ای مقایسه کردیم. این سه مجموعه داده شامل مجموعه داده‌های مصنوعی [۲۲]، مجموعه داده رهگیری فعالیت انسان بوسیله تلفن همراه *actitracker* [۳۱] و مجموعه داده *Cifar10* [۳۲] است. ما برای هر مجموعه داده یک *CNN* ساده‌تر شبکه‌های *CNN* کم‌عمق‌تر و برای مجموعه داده‌های پیچیده‌تر شبکه‌های کانولووشنال پیچیده‌تر طراحی شده است.

جدول (۴). الگوریتم پیشنهادی برای آموزش شبکه *CNN* با استفاده از تابع ضرر وزن دار افزایشی

Input: training data D, initial model M, class list C and hyperparameter T

Output: trained classifier M compatible with imbalanced data

the i^{th} class weight is initialised with $w_i^0 = 1$ where $i=1, \dots, C$ and (the dataset has a total of C classes)

the i^{th} class balanced weight is $WB_i = \frac{\sum_{i=1}^C n_i}{C \times n_i}$ where n_i is number of i^{th} class samples

For $e=0$ to the number of epochs E:

$$\alpha = e^{-\frac{(e-E)}{T}}$$

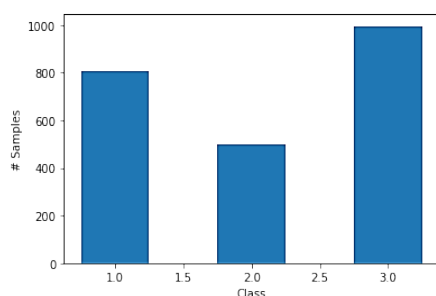
For $i=1$ to C:

$$w_i^e = \alpha \times BW_i + (1 - \alpha)w_i^0$$

w^e (normalizing (w^{e+1}))

Train M using D AND update model weight θ using IWCE loss function:

$$= \frac{1}{|D|} \sum_{(d_i, y_i) \in D_{new}} \sum_{c=1}^C w_c^e t_c^i \log p_c^i \quad // t^i \text{ is the target value and } p^i \text{ is the predicted probability vector using softmax on model M for } d_i$$



شکل (۳). توزیع داده‌های مصنوعی در سه دسته

همان‌طور که در *Error! Reference source not found.* نشان داده شده است، *CNN* پیشنهادی دارای شش لایه است که در انتهای لایه‌ی خروجی آن فعال‌ساز *soft-max* قرار گرفته است. لایه اول یک لایه کانولوشن یک‌بعدی با کرنل‌های 3×1 است و لایه دوم هم یک لایه کانولوشن یک‌بعدی با هسته‌های 2×1 است و به دنبال آن یک لایه 2×1 *max-pooling* قرار گرفته است. این دو لایه به ترتیب دارای تعداد ۳۲ و ۶۴ فیلتر هستند. پس از آنها، دو لایه کاملاً متصل به ترتیب دارای ۱۲۸ و ۱۶ نرون قرار دارد و *dropout* ۲۰ درصد در لایه ۱۲۸ نرون کاملاً متصل استفاده شده است. تمام لایه‌ها با استفاده از تابع *ReLU* فعال می‌شوند. در نهایت، لایه خروجی یک‌لایه کاملاً متصل با سه نرون است که با استفاده از تابع *softmax* فعال می‌شود. بهینه‌ساز *Adagard* با یادگیری برابر با ۰,۳ و *exponential scheduler* با گاما برابر با ۰,۹ استفاده شده است، وزن و بایاس در ابتدا با تنظیمات پیش‌فرض *PyTorch* استفاده شده است. لازم به ذکر است تعداد لایه‌ها و نحوه قرارگیری آنها با سعی و خطا و بررسی پیکربندی‌های مختلف و مقایسه دقت نهایی آنها انتخاب شده است.

جدول (۵). پیکربندی *CNN* استفاده شده برای مجموعه داده

داده‌های مصنوعی

لایه‌های شبکه <i>CNN</i> پیشنهادی:
1D Convolution 32 filters, 3x1 kernel and ReLU
1D Convolution 64 filters, 2x1 kernel
Max-Pooling 2x1 kernel
Fully connected 128 Neurons, ReLU
Dropout 20%
Fully connected 16 Neurons, ReLU
Fully connected 3 Neurons, SoftMax

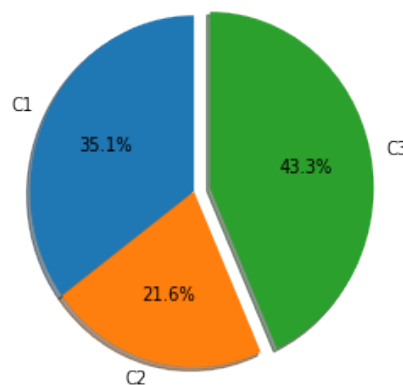
۱-۶. مجموعه داده مصنوعی

برای ایجاد مجموعه داده نامتوازن مصنوعی از توزیع نرمال استاندارد چندبعدی استفاده شده است. سه دسته در این مجموعه داده مصنوعی استفاده شده است که شامل ۱۲۳۰۰ نمونه است و هر نمونه یک بردار ۱۰ بعدی است. داده‌ها به شکلی که در [۲۳] بیان شده ایجاد شده است. رابطه زیر این داده‌ها را توصیف می‌کند. c مشخص‌کننده برچسب دسته است.

$$c = \begin{cases} 1. & 0 \leq \sum x_j^2 \leq X_{1/3}^2 \\ 2. & X_{1/3}^2 \leq \sum x_j^2 \leq X_{2/3}^2 \\ 3. & X_{2/3}^2 \leq \sum x_j^2 \end{cases} \quad (17)$$

در رابطه (۱۷) $X_{k/3}^2$ چندک^۱هایی هستند که مشخص‌کننده $(k/3)$ از کل داده‌های ۱۰ بعدی تولید شده با توزیع X^2 هستند و $\sum x_j^2$ فاصله اقلیدسی است. در واقع رابطه (۱۷) نشان دهنده داده‌های مربوط به دسته‌های مختلف است که توسط کره‌های چند بعدی متحدالمرکز تودرتو جدا می‌شوند. همانطور که در رابطه (۱۷) نشان داده شده است، نمونه‌های دسته ۱ در اطراف مبدا یک کره قرار گرفته‌اند، نمونه‌های دسته ۲ بین سطح دو کره توزیع شده‌اند و نمونه‌های دسته ۳ خارج از کره‌ها قرار دارند.

برای ایجاد یک مجموعه داده آموزشی نامتوازن دسته‌ای، در مجموع ۲۳۰۰ نمونه به‌عنوان داده‌های آموزشی استخراج می‌شود، به طوری که ۸۰۰، ۵۰۰ و ۱۰۰۰ نمونه به ترتیب، متعلق به دسته ۱، ۲ و ۳ هستند (*Error! Reference source not found.*) مجموعه آزمون با استخراج ۱۰۰۰۰ نمونه مستقل ساخته می‌شود. هر نمونه در مجموعه آزمون متعلق به یکی از سه دسته با احتمال‌های مساوی است. تقریباً به تعداد مساوی داده در مجموعه آزمون برای سه دسته وجود دارد.



<i>SMOTE-CNN</i>	95.48	93.45
<i>Adaboost-CNN</i>	95.61	94.08
<i>Incremental Weighted Cross-Entropy</i>	96.16	94.8

دومین روش پرطرفدار برای رسیدگی به نامتوازنی دسته‌ای، روش بیش نمونه برداری مصنوعی دسته اقلیت [۳۳] (*SMOTE*) است، از ترکیب *SMOTE* و *CNN* که در [۳۴] معرفی شده استفاده شده است. نتیجه *SMOTE* روی داده‌های مصنوعی در ردیف ۳ جدول ۶ نشان داده شده است، روش ما دقت آموزش و آزمون بهتری را نسبت به *SMOTE* نشان داده است. به طور کلی استفاده از *SMOTE* دقت *CNN* استاندارد با تابع ضرر آنتروپی متقابل را در نمونه‌های آزمایشی بهبود بخشیده است. ما همچنین روش خود را با روش جدید آدابوست چند دسته‌ای مبتنی بر *CNN* مقایسه کردیم که در [۲۲] معرفی شده است، روش ما در هر دو مرحله آموزش و آزمون به ترتیب با مقدار ۰.۵۵ و ۰.۷۲ بالاتر، دقت بهتری دارد.

۲-۶. مجموعه داده‌ی تشخیص فعالیت‌های انسان

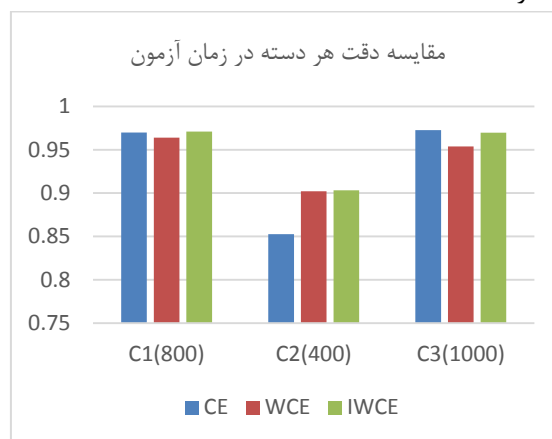
توسط گوشی هوشمند

برای انجام این کار از مجموعه داده *Actitracke* [۳۱] که به وسیله آزمایشگاه *WISDM* (*Wireless Sensor Data Mining*) جمع‌آوری شده است استفاده کرده‌ایم. این مجموعه داده شامل اطلاعات جمع‌آوری شده از ۳۶ فرد است که بوسیله گوشی همراهی که در جیبشان قرار داده شده و با سرعت ۲۰ نمونه بر ثانیه گرفته شده است. کاربران ۶ فعالیت مختلف را در محیط کنترل شده انجام می‌دهند و داده‌ها در واقع مقادیر شتاب سه بعدی در راستای محورهای x ، y و z است. این فعالیت‌ها شامل پایین رفتن از پله (*Downstairs*)، دویدن (*Jogging*)، بالا رفتن از پله‌ها (*Upstairs*)، ایستادن (*Standing*)، نشستن (*Sitting*) و پیاده روی (*Walking*) است.

دقت شبکه در زمان آزمون برای *CNN* پیشنهادی با تابع ضرر آنتروپی متقابل (*CE*)، تابع ضرر آنتروپی متقابل وزن دار (*WCE*) و تابع ضرر وزن دار افزایشی (*IWCE*) برای دسته‌های مختلف در مجموعه داده مصنوعی در شکل (۲) نشان داده شده است. روش پیشنهادی ما (*IWCE*) در دسته‌های اقلیت (*C2*) و متوسط (*C1*) بهترین نتایج را دارد و در دسته *C3* نتیجه قابل قبولی دارد که بهتر از *WCE* است.

دقت زمان آموزش و آزمون روی مجموعه داده مصنوعی در جدول (۱) نشان داده شده است. نتیجه نشان می‌دهد که دقت آموزش و آزمون تابع ضرر وزن دار افزایشی (*IWCE*) بالاتر از رویکردهای دیگر است، دقت آموزش ۱/۷۳ بیشتر از *CNN* با تابع ضرر آنتروپی متقابل استاندارد است و دقت آزمون در نمونه آزمایشی ۱۰۰۰۰ تایی هنگام استفاده از *IWCE* به جای تلفات متقابل آنتروپی استاندارد، ۱/۶۲ بیشتر است.

در مقایسه روش آدابوست مبتنی بر درخت تصمیم [۲۳] که یک راه حل موفق در مواجهه با نامتوازنی دسته‌ای است، دقت زمان آموزش و آزمون تابع ضرر پیشنهادی ما به ترتیب ۴/۳ و ۱۷/۷۷ بیشتر است.



شکل (۲) مقایسه دقت هر دسته با استفاده از سه تابع ضرر *CE*، *WCE* و *IWCE*

در مقایسه تابع ضرر پیشنهادی با آنتروپی متقابل وزن دار که یک راه حل رایج در مواجهه با نامتوازنی دسته‌ای است و هر دو به عنوان تابع ضرر *CNN* پیشنهادی در آموزش شبکه استفاده شده‌اند، دقت زمان آموزش و آزمون تابع ضرر پیشنهادی ما به ترتیب ۳۳/۰ و ۷/۰ بیشتر است جدول (۱) مقایسه نتایج تکنیک‌های مختلف روی داده‌های

مصنوعی

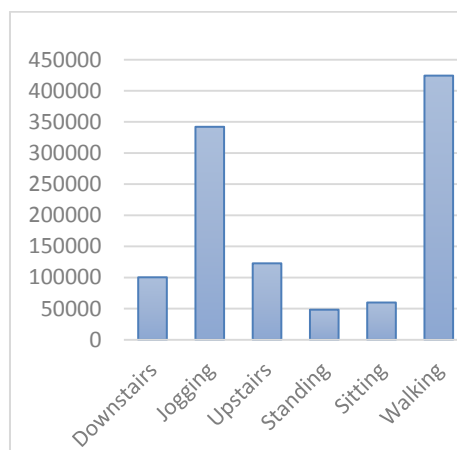
Model name	Train acc%	Test acc%
<i>AdaBoost-Decision-Tree</i>	91.78	77.08
<i>Cross-Entropy-CNN</i>	94.43	93.18
<i>Weighted -Cross-Entropy-CNN</i>	95.83	94.01

¹ Synthetic Minority Oversampling Technique

جدول (۷). ساختار *CNN* پیشنهادی برای مجموعه داده

actitracker		
layer	Number of filters/neurons	Kernel size
1D Convolution ReLU	100	10x1
Max-Pooling kernel	--	3x1
1D Convolution ReLU	160	10x1
Global-Average-Pooling Dropout 20%	--	--
Fully connected SoftMax	6 Neurons	--

استفاده شده است. ما *CNN* را به طور جداگانه با سه تابع ضرر مختلف آموزش داده ایم. تابع ضرر آنتروپی متقابل، تابع آنتروپی متقابل وزن دار و تابع ضرر پیشنهادی ما یعنی تابع ضرر وزن دار افزایشی. نتیجه در جدول جدول نشان داده شده است. با سه معیار مختلف یعنی دقت، دقت متوازن که میانگین دقت در همه دسته‌هاست و $F1$ -*score* نتایج بررسی شده‌اند.

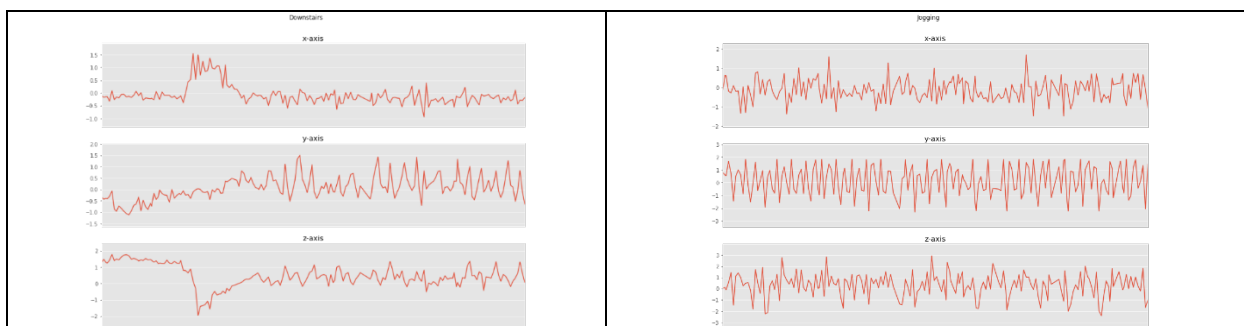


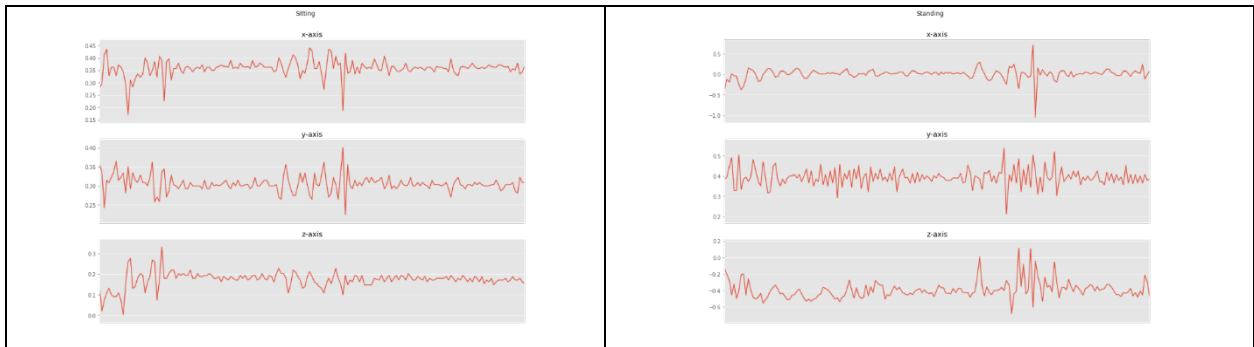
شکل (۶). توزیع دسته‌های مختلف در مجموعه داده

actitracker

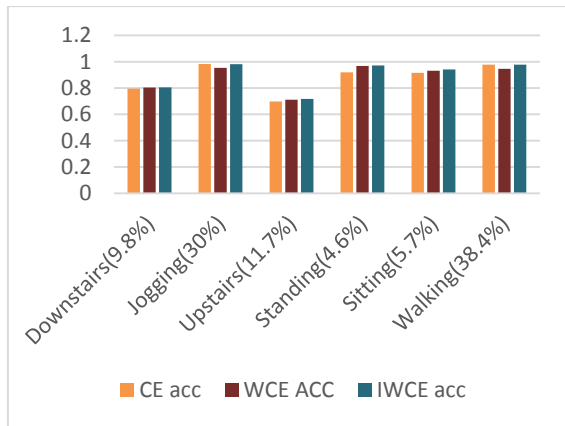
همان‌طور که در شکل نشان داده شده است، مجموعه داده نسبت به فعالیت‌های انجام شده (برچسب‌های دسته) نامتوازن است. در ابتدا داده‌ها نرمال‌سازی شده و به توزیع نرمال استاندارد برده شده است، سپس این داده‌های استاندارد شده به برش‌های زمانی با اندازه پنجره ۸۰ تقسیم می‌شوند که در واقع به مجموعه‌ای از قطعات ۴ ثانیه‌ای داده تبدیل می‌شود. سپس این داده‌ها به طور تصادفی با نسبت ۷۰ به ۳۰ به مجموعه‌های آموزشی و آزمون تقسیم می‌شوند. نمونه‌ای از داده‌ها در شکل نشان داده می‌شود، داده‌های آموزشی به یک شبکه کانولوشنال یک‌بعدی داده شد که ساختار شبکه در ادامه توضیح داده شده است.

شبکه *CNN* پیشنهادی با شش لایه در جدول نشان داده شده است، در ابتدا با یک لایه کانولوشنال یک بعدی با 1×10 فیلتر آغاز می‌شود و پس از آن یک لایه *Max-pooling* 3×1 قرار گرفته، یک لایه کانولوشنال یک بعدی دیگر با 160 فیلتر به عنوان لایه سوم قرار گرفته است و پس از *average-pooling* و *dropout* 20% درصد، لایه تماماً متصل با 6 نرون قرار داده ایم که یک تابع *SoftMax* در انتهای آن قرار داده شده است. بهینه‌سازی *Adam* با نرخ یادگیری 0.001 برای آموزش *CNN*





شکل (۷). نمونه‌های از داده‌های آموزشی نرمال شده.



شکل (۸). دقت هر یک از دسته‌ها در مجموعه داده **actitracker** با سه تابع ضرر **CE**، **WCE** و **IWCE**

۳-۶. مجموعه داده **Cifar10**

مجموعه داده **Cifar10** [۳۲] متوازن است و شامل ۱۰ دسته است که برای هر دسته ۵۰۰۰ داده آموزشی استفاده شده است. نمونه‌هایی از تصاویر هر دسته در نشان داده شده است و ۱۰۰۰۰ تصویر برای داده‌های آزمون ذخیره شده است. برای کار با این مجموعه داده‌ها، نیمی از دسته‌های این مجموعه داده، به صورت دستی با نسبت **IR (imbalance rate)** های مختلفی (۲، ۵، ۱۰، ۲۰، ۴۰ و ۶۰ درصد)، که در جدول ۹ با **p** مشخص شده است، نامتوازن شده است. در این مقاله دسته‌بند **Resnet18** را به عنوان شبکه عمیق پایه، استفاده کرده‌ایم. سه تابع ضرر مختلف را برای این دسته‌بند مورد آزمایش قرار داده‌ایم، تابع ضرر **CE**، تابع ضرر **WCE** و تابع ضرر پیشنهادی که **IWCE** نام دارد. سپس این دسته‌بندها را با مجموعه داده‌های مختلفی که از نامتوازن سازی مجموعه داده‌ی **CIFAR-10** بدست آمده، آموزش می دهیم، نتایج در جدول ۹ نشان داده شده است.

جدول (۸). مقایسه نتایج سه تابع ضرر **WCE**، **CE** و **IWCE** روی مجموعه داده **actitracker**

	Train accuracy	Test accuracy	Balanced_accuracy	F1-score
CNN+CE	93.53%	92.25%	88.14%	87.17%
CNN + weighted CE	92.62%	91.42%	88.86%	86.75%
CNN+incremental weighted CE	93.13%	92.92%	89.90%	87.89%

همان‌طور که از جدول مشخص است با در نظر گرفتن دقت آزمون، بافاصله ۰/۶۷ و ۱/۶ به ترتیب از آنتروپی متقابل و آنتروپی متقابل وزن دار بهتر است. با توجه به دقت متوازن و امتیاز **F1**، تابع ضرر وزن دار افزایشی به ترتیب با کسب ۸۹/۹۰ و ۸۷/۸۹ بهترین امتیاز را کسب کرده است. در شکل ۸ دقت دسته‌بند برای هر یک از دسته‌ها با سه تابع ضرر مختلف یعنی آنتروپی متقابل (**CE**)، آنتروپی متقابل وزن دار (**WCE**) و تابع ضرر وزن دار افزایشی (**IWCE**) محاسبه شده است.

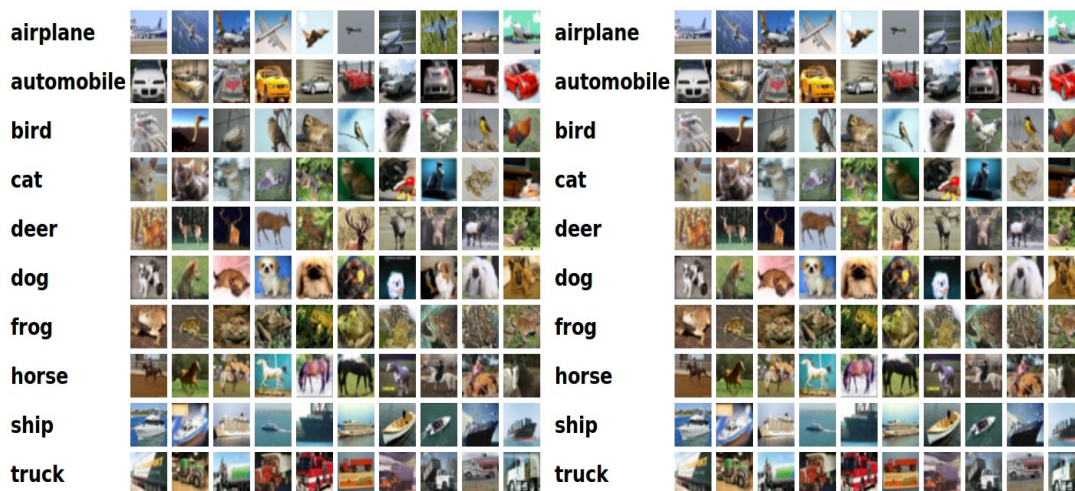
همان‌طور که از نتایج مشخص است تابع ضرر پیشنهادی ما توانسته است در دسته‌های اقلیت (**Standing** با ۴/۶ درصد داده و **sitting** با ۵/۷ درصد داده) بهترین نتیجه را کسب کرده است و در دیگر دسته دقت نزدیک به بهترین نتیجه بین **CE** و **WCE** را به دست آورده است.

دیگر از آن خودکرده است. شکل میانگین $F1$ -score در دسته‌های اقلیت برای سه تابع ضرر رسم شده است که نشان می‌دهد تابع ضرر آنتروپی متقابل وزن‌دار افزایشی نسبت به افزایش نامتوازنی، مقاومت بیشتری از خود نشان می‌دهد.

از معیارهای $F1$ -score و $Balanced-accuracy$ و میانگین $F1$ -score در دسته‌های اقلیت، برای بررسی کارایی الگوریتم خود استفاده کرده‌ایم، همان‌طور که از نتایج برمی‌آید، تابع ضرر $IWCE$ ، به‌جز در یک مورد، بهترین نتیجه را بین دو تابع ضرر

جدول (۹). مقایسه‌های معیارهای کارایی برای سه تابع ضرر CE ، WCE و IWE با میزان نامتوازنی دسته‌های مختلف

M=5		p=60	P=40	P=20	P=10	P=5	P=2
Balanced accuracy	CE loss function	0.6834	0.6432	0.5186	0.7126	0.6569	0.5327
	WCE loss function	0.8283	0.81235	0.7783	0.7099	0.6464	0.6068
	IWCE loss function	0.8386	0.82725	0.7865	0.7257	0.6923	0.6134
'F1-score'	CE loss function	0.84172	0.83366	0.83914	0.84258	0.85201	0.86114
	WCE loss function	0.83803	0.81496	0.81933	0.81174	0.80040	0.80483
	IWCE lossfunction	0.84876	0.84147	0.83787	0.84921	0.85631	0.85438
'F1-score' Minority avg	CE loss function	0.82462	0.7762	0.69278	0.55767	0.43952	0.21412
	WCE loss function	0.8146	0.75613	0.72032	0.54271	0.31115	0.25828
	IWCE loss function	0.8149	0.7921	0.7165	0.5935	0.5326	0.31646



شده سعی دارد درعین‌حال که اهمیت دسته اقلیت را مرحله‌به‌مرحله افزایش می‌دهد، بر خلاف تابع آنتروپی متقابل وزن‌دار، آموزش با داده‌های دسته اکثریت را هم از دست ندهد. آزمایش این تابع ضرر جدید روی سه مجموعه‌داده مختلف، یعنی داده‌های ۱۰ بعدی مصنوعی، داده‌های $actitracker$ و مجموعه‌داده‌های cifar10 انجام شده است. نتایج کارایی تابع ضرر پیشنهادی را در هر سه مجموعه‌داده نشان می‌دهد، همچنین نتایج نشان می‌دهد این تابع ضرر مقاومت بیشتری نسبت به افزایش درصد نامتوازنی از خود نشان می‌دهد. این تابع ضرر پیشنهادی روی مجموعه‌داده با پیچیده‌تر بررسی نشده است که

۷. نتیجه‌گیری و راهکارهای آینده

باتوجه‌به اینکه، در سال‌های اخیر مسئله‌ی دسته‌بندی داده‌های نامتوازن به عنوان یکی از چالش‌های اصلی در حوزه داده‌کاوی، مورد توجه بسیاری از محققین قرار گرفته است، در این پژوهش، به این مسئله پرداخته شده و تلاش شده است تا با به‌کاربردن یکی از موفق‌ترین مدل‌های یادگیری ماشین یعنی شبکه‌های عصبی کانولوشنال و با معرفی تابع ضرر آنتروپی متقابل وزن‌دار افزایشی که یک تابع ضرر به نفع دسته اقلیت است، مشکلات مسئله دسته‌بندی نامتوازن را به حداقل برسانیم. تابع ضرر معرفی

- [14] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," in 2017 IEEE International Conference on Computer Vision (ICCV), 22-29 Oct. 2017 2017, pp. 2999-3007, doi: 10.1109/ICCV.2017.324.
- [15] H. Wang, Z. Cui, Y. Chen, M. Avidan, A. B. Abdallah, and A. Kronzer, "Predicting Hospital Readmission via Cost-Sensitive Deep Learning," IEEE/ACM Transactions on Computational Biology and Bioinformatics, Article vol. 15, no. 6, pp. 1968-1978, 11/1 2018, doi: 10.1109/TCBB.2018.2827029.
- [16] S. H. Khan, M. Hayat, M. Bennamoun, F. A. Sohel, and R. Togneri, "Cost-Sensitive Learning of Deep Feature Representations From Imbalanced Data," IEEE Trans. Neural Networks Learn. Syst., vol. 29, no. 8, pp. 3573-3587, 2018, doi: 10.1109/TNNLS.2017.2732482.
- [17] C. Zhang, K. C. Tan, and R. Ren, "Training cost-sensitive Deep Belief Networks on imbalance data problems," in 2016 International Joint Conference on Neural Networks (IJCNN), 24-29 July 2016 2016, pp. 4362-4367, doi: 10.1109/IJCNN.2016.7727769.
- [18] Y. Zhang, L. Shuai, Y. Ren, and H. Chen, "Image classification with category centers in class imbalance situation," in 2018 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC), 18-20 May 2018 2018, pp. 359-363, doi: 10.1109/YAC.2018.8406400.
- [19] W. Ding, D. Huang, Z. Chen, X. Yu, and W. Lin, "Facial action recognition using very deep networks for highly imbalanced class distribution," in 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 12-15 Dec. 2017 2017, pp. 1368-1372, doi: 10.1109/APSIPA.2017.8282246.
- [20] S. Ando and C. Y. Huang, "Deep Over-sampling Framework for Classifying Imbalanced Data," in Machine Learning and Knowledge Discovery in Databases, Cham, M. Ceci, J. Hollmén, L. Todorovski, C. Vens, and S. Džeroski, Eds., 2017// 2017: Springer International Publishing, pp. 770-785.
- [21] A. Y. A. Saeed and A. E. B. Alawi, "Covid-19 Diagnosis Model Using Deep Learning with Focal Loss Technique," in 2021 International Congress of Advanced Technology and Engineering (ICOTEN), 4-5 July 2021 2021, pp. 1-4, doi: 10.1109/ICOTEN52080.2021.9493477.
- [22] A. Taherkhani, G. Cosma, and T. M. McGinnity, "AdaBoost-CNN: An adaptive boosting algorithm for convolutional neural networks to classify multi-class imbalanced datasets using transfer learning," Neurocomputing, vol. 404, pp. 351-366, 2020/09/03/ 2020, doi: https://doi.org/10.1016/j.neucom.2020.03.064.
- [23] T. J. Hastie, S. Rosset, J. Zhu, and H. Zou, "Multi-class AdaBoostB " Statistics and Its Interface, vol. 2, pp. 349-360, 2009.
- [24] O. EL ZEIN, M. M. SOLIMAN, A. ELKHOLY, and N. I. GHALI, "MULTI-CLASSIFICATION MODEL FOR COVID-19 PREDICTION USING IMBALANCED X-RAY DATASET BASED ON TRANSFER LEARNING AND CLASS WEIGHTING-SMOTE METHOD," Journal of Theoretical and Applied Information Technology, vol. 100, no. 5, 2022.
- [25] م. سبک رور "تشخیص و مکان‌یابی and م. خالونی, م. فخر دانش "رویدادهای رایج و نادر در ویدیو با بکارگیری شبکه تخصصی مولد" (in persian), vol. 8, no. 3, pp. 40-51, 2019. [Online]. Available: http://jscit.nit.ac.ir/article_93041.html
http://jscit.nit.ac.ir/article_93041_98448d599e92c5238380a9006d717889.pdf.
- [26] A. Choromanska, M. Henaff, M. Mathieu, G. B. Arous, and Y. LeCun, "The loss surfaces of multilayer

به‌عنوان راهکارهای آتی پیشنهاد می‌شود و همچنین ترکیب این تابع ضرر با دیگر تکنیک‌های مقابله با نامتوازنی دسته‌ای می‌تواند در آینده بررسی شود.

مراجع

- [1] س. ع. هاشمی and م. خسروی, ح. شیرازی, ک. دانش‌تبار احمدی گلپایگانی, "تشخیص شایعه در شبکه‌های اجتماعی مبتنی بر تحلیل "الگوی فراوانی درجه نروس در زیرگراف های گام به گام انتشار" (in persian), پیافند الکترونیکی و سایبری, vol. 10, no. 3, pp. 93-105, 2022. [Online]. Available: https://ecdj.ihu.ac.ir/article_207814.html
https://ecdj.ihu.ac.ir/article_207814_3b093c167532561d02691da4758fc50f.pdf.
- [2] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," Journal of Big Data, vol. 6, no. 1, p. 27, 2019/03/19 2019, doi: 10.1186/s40537-019-0192-5.
- [3] G. Aguiar, B. Krawczyk, and A. Cano, "A survey on learning from imbalanced data streams: taxonomy, challenges, empirical study, and reproducible experimental framework," 2022. [Online]. Available: http://arXiv.org/abs/.
- [4] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Learning Deep Representation for Imbalanced Classification," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 27-30 June 2016 2016, pp. 5375-5384, doi: 10.1109/CVPR.2016.580.
- [5] W. Zhang, X. Li, X.-D. Jia, H. Ma, Z. Luo, and X. Li, "Machinery fault diagnosis with imbalanced data using deep generative adversarial networks," Measurement, vol. 152, p. 107377, 2020/02/01/ 2020, doi: https://doi.org/10.1016/j.measurement.2019.107377.
- [6] Q. Dong, S. Gong, and X. Zhu, "Imbalanced Deep Learning by Minority Class Incremental Rectification," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 6, pp. 1367-1381, 2019, doi: 10.1109/tpami.2018.2832629.
- [7] R. Anand, K. G. Mehrotra, C. K. Mohan, and S. Ranka, "An improved algorithm for neural network classification of imbalanced training sets," IEEE Transactions on Neural Networks, vol. 4, no. 6, pp. 962-969, 1993, doi: 10.1109/72.286891.
- [8] D. Masko and P. Hensman, "The Impact of Imbalanced Training Data for Convolutional Neural Networks," 2015.
- [9] H. Lee, M. Park, and J. Kim, "Plankton classification on imbalanced large scale database via convolutional neural networks with transfer learning," 2016 IEEE International Conference on Image Processing (ICIP), pp. 3713-3717, 2016.
- [10] S. Pouyanfar et al., "Dynamic Sampling in Convolutional Neural Networks for Imbalanced Data Classification," in 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 10-12 April 2018 2018, pp. 112-117, doi: 10.1109/MIPR.2018.00027.
- [11] M. Havaei et al., "Brain tumor segmentation with Deep Neural Networks," Medical Image Analysis, vol. 35, pp. 18-31, 2017/01/01/ 2017, doi: https://doi.org/10.1016/j.media.2016.05.004.
- [12] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," Neural Networks, vol. 106, pp. 249-259, 2018/10/01/ 2018, doi: https://doi.org/10.1016/j.neunet.2018.07.011.
- [13] S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng, and P. J. Kennedy, "Training deep neural networks on imbalanced data sets," ed. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE), 2016, pp. 4368-4374.

- [30] P. Shamsolmoali, M. Zareapoor, L. Shen, A. H. Sadka, and J. Yang, "Imbalanced data learning by minority class augmentation using capsule adversarial networks," *Neurocomputing*, 2020/07/28/ 2020, doi: <https://doi.org/10.1016/j.neucom.2020.01.119>.
- [31] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity Recognition Using Cell Phone Accelerometers," *SIGKDD Explor. Newsl.*, vol. 12, no. 2, pp. 74-82, 2011, doi: 10.1145/1964897.1964918.
- [32] A. Krizhevsky, "Learning Multiple Layers of Features from Tiny Images," University of Toronto, 05/08 2012.
- [33] N. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res. (JAIR)*, vol. 16, pp. 321-357, 01/01 2002, doi: 10.1613/jair.953.
- [34] J. H. Joloudari, A. Marefat, M. A. Nematollahi, S. S. Oyelere, and S. Hussain, "Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks," *Applied Sciences*, vol. 13, no. 6, doi: 10.3390/app13064006.
- networks," *Journal of Machine Learning Research, Conference article* vol. 38, pp. 192-204, 2015. [Online]. Available: <http://www.scopus.com/inward/record.url?scp=84954310140&partnerID=8YFLogxK>
- <http://www.scopus.com/inward/citedby.url?scp=84954310140&partnerID=8YFLogxK>.
- [27] K. Kawaguchi, Deep Learning without Poor Local Minima, S. a. D. Lee and U. L. , I. Guyon , R. Garnett, eds.: Curran Associates, Inc., 2016. [Online]. Available: <https://proceedings.neurips.cc/paper/2016/file/f2fc990265c712c49d51a18a32b39f0c-Paper.pdf>.
- [28] B. Guo, S. Chen, Z. Hong, and G. Xu, "Pattern Recognition and Analysis: Neural Network using Weighted Cross Entropy," *Journal of Physics: Conference Series*, vol. 2218, no. 1, p. 012043, 2022/03/01 2022, doi: 10.1088/1742-6596/2218/1/012043.
- [29] Z. Zhou, H. Huang, and B. Fang, "Application of weighted cross-entropy loss function in intrusion detection," *Journal of Computer and Communications*, vol. 9, no. 11, pp. 1-21, 2021.