



پیش بینی زباله تولیدی تهران با استفاده از شبکه عصبی مصنوعی و روش های آماری چند متغیره

محمد علی عبدلی

استاد دانشکده محیط زیست دانشگاه تهران

روح اله نوری

دانشجوی کارشناسی ارشد عمران-محیط زیست

دانشگاه تربیت مدرس

مهدی جلیلی

دانشجوی دکتری محیط -زیست دانشگاه تهران

احسان صالحیان

دانشجوی کارشناسی ارشد عمران محیط زیست،

دانشگاه تربیت مدرس

۱- مقدمه

مواد زائد جامد شهری نتیجه طبیعی فعالیت های انسان می باشد. در صورتی که سیستم مدیریت مناسبی برای چنین امری به کار گرفته نشود، این مواد باعث آلودگی های زیستی - محیطی زیادی می شوند و سلامت بشر را به خطر می اندازند. ایجاد چنین سیستمی به خاطر پیچیدگی و طبیعت بسیار ناهمگن تولید زباله، کاری بسیار مشکل می باشد. یکی از عوامل بسیار مهم در راه اندازی صحیح چنین سیستم پیچیده ای شناخت کمیت زباله تولیدی در آن است، زیرا کمیت تولید در حجم سرمایه گذاری برای ماشین آلات، ظروف ذخیره در محل، ایستگاه های انتقال، ظرفیت دفع، سازماندهی و تشکیلات مناسب موثر است. در تخمین میزان زباله تولیدی برای یک شهر داشتن الگوهای فصلی زباله تولیدی می تواند نقش موثری داشته باشد، به همین منظور برای پیش بینی زباله تولیدی در شهر تهران اقدام به ساخت سری زمانی زباله تولیدی این شهر گردید به طوری که مقدار زباله در هفته t تابعی از مقدار زباله در هفته $t-1, t-2, \dots, t-12$ قرار داده شد.

روش های سنتی برای تخمین میزان تولیدی جامدات زائد یک جامعه غالباً بر اساس فاکتورهای نظیر آمار جمعیت و فاکتورهای اقتصادی-اجتماعی آن جامعه استوار می باشد. که به صورت ضرایب تولید به ازای هر نفر محاسبه می شود. اشکال این روش در این است که این ضرایب ممکن است با زمان تغییر کند و برای یک سیستم مدیریت مواد زائد جامد به لحاظ دینامیک بودن آن کارایی خود را از دست دهد. به همین دلیل آرایه الگوهای نو و به کارگیری تکنیک های پیشرفته می تواند در برآورد این سیستم دینامیک و غیرخطی موثر باشد. این روش ها عمدتاً شامل استفاده از

چکیده

پایه و اساس برنامه ریزی و طراحی سیستم مدیریت مواد زائد جامد شهری، شناخت کمیت و کیفیت تولید است. تخمین میزان زباله تولیدی به دلیل نوسانات زیاد تولید و پارامترهای کوانتونی که بر آن موثر است یکی از کارهای بسیار دشوار در امر مدیریت مواد زائد جامد می باشد. در این مطالعه با استفاده از شبکه عصبی مصنوعی مدلی مناسب برای تخمین وزن زباله تولیدی در شهر تهران ارائه گردیده است و نتایج آن با مدل ترکیبی رگرسیون خطی چندمتغیره و آنالیز مولفه های اصلی مقایسه شده است. برای این منظور از سری زمانی زباله تولیدی شهر تهران در فاصله زمانی ۱۳۸۲ تا سه ماهه نخست ۱۳۸۵ که به صورت هفتگی مرتب شده بودند استفاده گردید. نتایج به دست آمده حاکی از برتری مطلق نتایج شبکه عصبی در مقایسه با مدل ترکیبی رگرسیونی و آنالیز مولفه های اصلی می باشد. به طوریکه ضریب همبستگی در مدل شبکه عصبی ارائه شده، برای مرحله تست شبکه معادل ۰/۸۴۷ بود.

واژگان کلیدی: زباله تولیدی، شبکه عصبی مصنوعی،

آنالیز مولفه های اصلی، رگرسیون خطی چندمتغیره، تهران



آموزش و صحت یابی شبکه از یکدیگر فاصله بگیرند. برای رفع این مشکل میتوان از روش STA (روش توقف آموزش) استفاده نمود. در این روش داده‌ها به سه بخش تقسیم می‌شوند. بخش اول مربوط به آموزش شبکه، بخش دوم جهت توقف محاسبات هنگامی که خطای صحت‌یابی شروع به افزایش می‌نماید، و بخش سوم داده‌ها که برای صحت‌یابی شبکه استفاده می‌گردد. معماریهای مختلفی برای STA آزمایش گردید و در نهایت شبکه با سه لایه که لایه اول دارای ۱۳، لایه دوم ۲۲ و لایه سوم ۱ نورون بودند به عنوان ساختار برتر شبکه انتخاب شد.

۲-۳- آنالیز مولفه اصلی

PCA یکی از روش‌های آماری چندمتغیره می‌باشد که می‌توان از آن برای حذف همبستگی بین متغیرهای ورودی به مدل رگرسیونی و تفسیر راحت‌تر متغیرها استفاده نمود. با اعمال PCA متغیرهای ورودی اصلی به متغیرهای جدید که بدون همبستگی می‌باشند تبدیل می‌شوند. مولفه‌های ایجاد شده ترکیبی خطی از متغیرهای اصلی می‌باشند. (Lu et al., 2003) این آنالیز به عنوان یکی از روش‌های چندمتغیره‌ای است که کار آن یافتن ترکیباتی از P متغیر X_1, X_2, \dots, X_p جهت ایجاد P مولفه مستقل Z_1, Z_2, \dots, Z_p می‌باشد. عدم همبستگی بین این مولفه‌ها یک ویژگی مفید است زیرا عدم همبستگی به این معنی است که مولفه‌ها جنبه‌های متفاوتی از پارامترهای اصلی را نمایان می‌سازند. در این روش اطلاعات پارامترهای اصلی با کمترین تلفات در مولفه‌های حاصل آورده می‌شود (Helena et al., 2000). هر مولفه اصلی میتواند با دنباله زیر مشخص شود:

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$$

که Z_i معرف مولفه موردنظر، a_i بردار ویژه مربوطه و X نیز متغیرهای اصلی می‌باشد. این اطلاعات از حل معادله زیر به دست می‌آید.

$$|R - \lambda I| = 0$$

که در آن I ماتریس واحد، R ماتریس واریانس-کوواریانس و λ نیز مقادیر ویژه می‌باشد. از این مقادیر ویژه، بردارهای ویژه به دست می‌آیند. برای مشخص کردن

مدل‌ها، روش‌های آماری کلاسیک و جدید تکنیک‌های نو مانند منطق فازی، مدل‌های سری زمانی و شبکه‌های عصبی مصنوعی می‌باشند.

اخیرا استفاده از شبکه‌های عصبی مصنوعی و روش‌های آماری چند متغیره در موضوعات مهم زیست محیطی مانند آلودگی هوا (Sahin et al., 2005; Lu et al., 2004)، آلودگی آب‌های سطحی (Sahoo et al., 2006; Shrestha, S. and Kazama, F., 2007) و به تبع آن در امر مدیریت مواد زائد جامد نیز رواج یافته است، که در این مورد می‌توان به ارایه مدلی بر اساس شبکه عصبی برای کنترل نرخ جریان شیرابه در محل دفن مواد زائد جامد شهری در استانبول ترکیه (Karaca, F. and Özkaya, B. 2006)، تخمین مقدار تولید مواد زائد جامد و فاکتور تولید با استفاده از شبکه عصبی مصنوعی و پیش‌بینی میزان گرمای تولیدی از مواد زائد جامد شهری با استفاده از شبکه عصبی و رگرسیون خطی چندمتغیره در شهر نانجینگ چین (Dong et al., 2003) اشاره کرد. در این مطالعه از دو روش شبکه عصبی Feed-Forward و رگرسیون خطی چندمتغیره و ارایه یک روش جدید از کاربرد PCA برای پردازش داده‌های ورودی به مدل رگرسیون خطی چندمتغیره جهت تخمین میزان زباله تولیدی هفتگی شهر تهران استفاده شده است. اهداف این مطالعه عبارتند از: (۱) تخمین میزان زباله تولیدی با استفاده از مدل رگرسیون خطی چندمتغیره و PCA، (۲) برآورد میزان زباله تولیدی با استفاده از شبکه عصبی Feed-Forward و (۳) مقایسه نتایج این دو مدل و انتخاب مدل مناسب.

۲- مواد و روش‌ها

۲-۱- منطقه مورد مطالعه و داده‌های مساله

۲-۲- شبکه عصبی مصنوعی

مدل‌های شبکه عصبی الگو گرفته از مغز انسان می‌باشند. استفاده از این مدل در موضوعات زیست محیطی از دهه ۱۹۴۰ شروع شده و از ۱۹۹۰ به بعد بطور وسیعی مورد استفاده قرار گرفته است. مشکلی که در شبکه عصبی وجود دارد مشکل فوق برازشی است که در این حالت شبکه آموزش دچار اشکال گردیده و باعث می‌شود که نتایج



اهمیت ترین آنها صورت میگیرد. معیار میزان اهمیت متغیر در مدل، مقدار سطح معنی داری یا آماره F متناظر با آن در جدول آزمون معنی داری متغیرهاست.

۲-۴- معیار ارزیابی اعتبار مدل‌های مورد

استفاده

جهت بررسی اعتبار مدل‌های رگرسیونی و شبکه عصبی از معیارهای برازندگی ضریب همبستگی (R) و میانگین نسبی خطای مطلق (MARE) استفاده گردید که هر کدام از این پارامترها در زیر تعریف شده اند. ضریب همبستگی بخشی از واریانس موجود در داده های مشاهده شده که توسط مدل قابل توجیه می باشد را توصیف می کند. محدوده تغییرات این پارامتر بین صفر و یک بوده و مقادیر بالاتر آن تطابق بهتر داده های مشاهده ای و برآورد شده را نشان می دهد. مطالعات Legates and McCabe (1999) نشان داد مقادیر R2 تحت تاثیر داده های پرت می باشد و باید آن را به اتفاق پارامترهای دیگر استفاده نمود. به همین دلیل در این تحقیق از معیار میانگین نسبی خطای مطلق (MARE) نیز استفاده شده است.

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - O')^2}}$$

$$MARE = \frac{1}{n} \sum_{i=1}^n \frac{|O_i - P_i|}{O_i}$$

۳- بحث و نتایج

۳-۱- آنالیز مولفه های اصلی

بررسی اولیه نشان داد که بین متغیرهای ورودی مورد استفاده در این مطالعه همبستگی معنی داری وجود دارد که برای از بین بردن این مشکل از روش PCA استفاده گردید. مقدار $KMO=0.704$ نیز امکان اجرای PCA را تأیید کرد. برای اجرای PCA پس از استاندارد کردن داده ها ماتریس مربعی واریانس- کوواریانس از مرتبه ۱۳ (معادل با تعداد متغیرهای ورودی) تشکیل شد، که نتایج آن در زیر آمده است.

امکان اجرای PCA بر روی پارامترهای ورودی نیاز به محاسبه فاکتور KMO میباشد. در صورتی که این فاکتور بزرگتر از ۵ به دست آید، نشان دهنده امکان اجرای PCA بر داده های اصلی میباشد.

۲-۳- رگرسیون خطی چند متغیره

مدل رگرسیونی به فرم ماتریسی را میتوان به صورت معادله زیر نشان داد:

$$Y = X\beta + e$$

در معادله بالا β ماتریس ضرایب رگرسیون e ماتریس خطای برازش و Y نیز ماتریس پاسخ می باشد. با حل معادله بالا بر حسب β خواهیم داشت:

$$\beta = (X'X)^{-1}(X'Y)$$

برای محاسبه معکوس ($X'X$) نباید بین متغیرهای مستقل همبستگی زیادی وجود داشته باشد زیرا در این صورت ماتریس ($X'X$) را نمی توان معکوس کرد و باعث افزایش خطا در اثر گرد کردن داده ها و محاسبات می شود. برای رفع این مشکل باید قبل از ساخت مدل رگرسیونی، همبستگی بین متغیرهای مستقل را به طریقی از بین برد. یک روش برای این کار استفاده از آنالیز اجزای اصلی (PCA) بر روی متغیرهای مستقل ورودی به مدل می باشد. معیار قضاوت برای رفع این مشکل با اجرای PCA بر روی داده های اصلی، میزان تحمل و فاکتور تورم واریانس می باشد. مقدار تحمل، نسبتی از تغییرات یک متغیر است که به وسیله سایر متغیرهای مستقل بیان نمی شود و مقدار آن بین صفر و یک تغییر می کند. به طور کلی هرچه مقدار تحمل بیشتر باشد، متغیر راحت تر وارد مدل می شود. عدد ایده آل برای مقدار تحمل ۱ می باشد، این مقدار برای تورم واریانس نیز مینیمم مقدار آن یعنی یک می باشد و مقادیر بزرگتر از ۵ برای فاکتور تورم واریانس بیانگر وجود مشکل در لحاظ آماری برای مدل رگرسیونی می باشد (Hocking, 2003).

در این تحقیق پس از رفع مشکل همبستگی در متغیرهای مستقل، مدلی مناسب با استفاده از تکنیک رگرسیون چند متغیره برای پیش بینی وزن زباله تولیدی توسعه یافت. در محاسبات رگرسیونی از الگوریتم گام به گام (Stepwise) استفاده گردید. در این روش ورود متغیرها به مدل رگرسیون به صورت مرحله ای از مهمترین متغیر تا کم



1.00	-0.02	0.09	0.16	0.19	0.04	-0.09	-0.08	-0.15	-0.03	0.12	0.38	0.63
-0.02	1.00	0.58	0.27	0.05	0.02	-0.02	-0.08	-0.12	-0.11	-0.04	0.00	0.01
0.09	0.58	1.00	0.57	0.23	0.04	-0.02	-0.07	-0.12	-0.13	-0.10	-0.01	0.03
0.16	0.27	0.57	1.00	0.56	0.22	0.00	-0.05	-0.08	-0.12	-0.12	-0.08	-0.01
0.19	0.05	0.23	0.56	1.00	0.56	0.22	0.01	-0.06	-0.05	-0.09	-0.07	-0.04
0.04	0.02	0.04	0.22	0.56	1.00	0.55	0.24	0.02	-0.01	-0.03	-0.09	-0.10
-0.09	-0.02	-0.02	0.00	0.22	0.55	1.00	0.58	0.25	0.08	0.02	-0.01	-0.10
-0.08	-0.08	-0.07	-0.05	0.01	0.24	0.58	1.00	0.60	0.30	0.12	0.05	0.01
-0.15	-0.12	-0.12	-0.08	-0.06	0.02	0.25	0.60	1.00	0.62	0.33	0.15	0.03
-0.03	-0.11	-0.13	-0.12	-0.05	-0.01	0.08	0.30	0.62	1.00	0.62	0.34	0.15
0.12	-0.04	-0.10	-0.12	-0.09	-0.03	0.02	0.12	0.33	0.62	1.00	0.61	0.32
0.38	0.00	-0.01	-0.08	-0.07	-0.09	-0.01	0.05	0.15	0.34	0.61	1.00	0.61
0.63	0.01	0.03	-0.01	-0.04	-0.10	-0.10	0.01	0.03	0.15	0.32	0.61	1.00

از حل این ماتریس ۱۳ مقدار ویژه و به ازای هر مقدار ویژه ۱۳ بردار ویژه حاصل شد که با استفاده از آنها، ۱۳ مولفه از متغیرهای اولیه ایجاد گردید. چون در بسیاری از موارد تعدادی از متغیرها به بیش از یک مولفه بستگی دارند، تعبیر مولفه‌ها مشکل خواهد بود. از این رو با استفاده از روش دوران مولفه‌ها می‌توان بدون تغییر میزان اشتراک، تفسیر مولفه‌ها را ساده‌تر کرد. در این تحقیق برای امکان تفسیر مولفه‌های ایجاد شده از چرخش استاندارد Varimax استفاده گردید ولی برای محاسبات رگرسیونی از مولفه‌های به دست آمده بدون چرخش استفاده شد. مشخصات هر مولفه که بدون چرخش و با چرخش Varimax ایجاد شده اند در زیر آمده است.

Component Matrix

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
VAR00001	0.15	0.06	0.71	-0.44	0.15	-0.27	0.21	0.26	0.18	-0.07	-0.16	0.10	-0.01
VAR00002	-0.31	0.22	0.35	0.61	0.34	0.30	0.32	-0.15	0.19	-0.01	0.05	0.04	-0.03
VAR00003	-0.38	0.38	0.47	0.54	0.13	-0.07	-0.13	0.21	-0.27	-0.06	-0.15	-0.12	0.06
VAR00004	-0.39	0.55	0.41	0.19	-0.30	-0.27	-0.25	0.00	0.12	0.17	0.16	0.17	-0.09
VAR00005	-0.26	0.69	0.22	-0.28	-0.42	-0.01	0.08	-0.21	0.06	-0.24	0.00	-0.18	0.09
VAR00006	-0.07	0.75	-0.11	-0.39	-0.04	0.34	0.18	0.01	-0.17	0.27	-0.13	0.04	-0.08
VAR00007	0.19	0.68	-0.34	-0.19	0.39	0.20	-0.16	0.15	0.00	-0.20	0.18	0.15	0.09
VAR00008	0.47	0.52	-0.37	0.08	0.37	-0.30	-0.10	-0.03	0.17	0.10	-0.06	-0.25	-0.13
VAR00009	0.66	0.30	-0.31	0.32	-0.08	-0.35	0.14	-0.17	-0.08	0.01	-0.14	0.20	0.19
VAR00010	0.76	0.15	-0.04	0.30	-0.36	0.01	0.23	0.19	-0.10	-0.11	0.15	-0.03	-0.22
VAR00011	0.74	0.02	0.26	0.17	-0.25	0.36	-0.13	0.15	0.19	0.15	-0.02	-0.09	0.22
VAR00012	0.64	-0.04	0.56	-0.04	0.08	0.23	-0.28	-0.24	-0.04	-0.12	-0.16	0.09	-0.18
VAR00013	0.44	-0.07	0.69	-0.23	0.27	-0.15	0.09	-0.14	-0.20	0.12	0.26	-0.09	0.09

Rotated Component Matrix

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
VAR00001	0.93	-0.02	0.04	0.02	-0.03	-0.04	0.07	-0.02	0.09	-0.07	0.15	0.03	0.28
VAR00002	-0.02	0.96	-0.01	0.01	-0.03	0.00	0.10	-0.04	0.00	-0.04	0.01	0.26	0.01
VAR00003	0.04	0.33	-0.04	-0.01	-0.02	0.00	0.28	-0.04	0.08	-0.04	0.00	0.89	0.02
VAR00004	0.07	0.12	-0.04	0.08	-0.01	-0.02	0.90	-0.04	0.28	-0.02	-0.04	0.27	-0.01
VAR00005	0.10	0.00	-0.04	0.29	-0.02	0.09	0.28	0.00	0.90	-0.02	-0.03	0.08	-0.02
VAR00006	0.02	0.01	0.00	0.91	0.09	0.28	0.08	-0.01	0.27	-0.01	-0.04	-0.01	-0.04
VAR00007	-0.04	0.00	0.01	0.28	0.29	0.91	-0.02	0.01	0.09	0.09	0.00	0.00	-0.04
VAR00008	-0.03	-0.04	0.03	0.09	0.90	0.30	-0.01	0.11	-0.02	0.29	0.01	-0.02	0.01
VAR00009	-0.08	-0.05	0.12	-0.01	0.31	0.09	-0.02	0.30	-0.02	0.88	0.05	-0.04	0.01
VAR00010	-0.02	-0.05	0.30	-0.01	0.12	0.01	-0.04	0.88	0.00	0.30	0.12	-0.05	0.04
VAR00011	0.04	-0.01	0.89	0.00	0.03	0.01	-0.04	0.30	-0.04	0.12	0.29	-0.04	0.11
VAR00012	0.18	0.01	0.30	-0.04	0.01	0.01	-0.04	0.13	-0.03	0.05	0.88	0.00	0.28
VAR00013	0.37	0.01	0.12	-0.04	0.01	-0.04	-0.01	0.05	-0.02	0.01	0.28	0.02	0.87



۳-۲- رگرسیون خطی چندمتغیره

پس از رفع مشکل همبستگی در متغیرهای اصلی، مدلی مناسب با استفاده از تکنیک رگرسیون چند متغیره با الگوریتم گام به گام (Stepwise) برای پیش بینی وزن زباله تولیدی توسعه یافت. که نتایج آن در جدول زیر آمده است.

Input Component	R Square	Sig. F Change	Tolerance	VIF
PC3	0.332	0.001	0.938	1.066
PC3,PC2	0.412	0.010	0.942	1.061
PC3,PC2,PC9	0.467	0.015	0.990	1.010

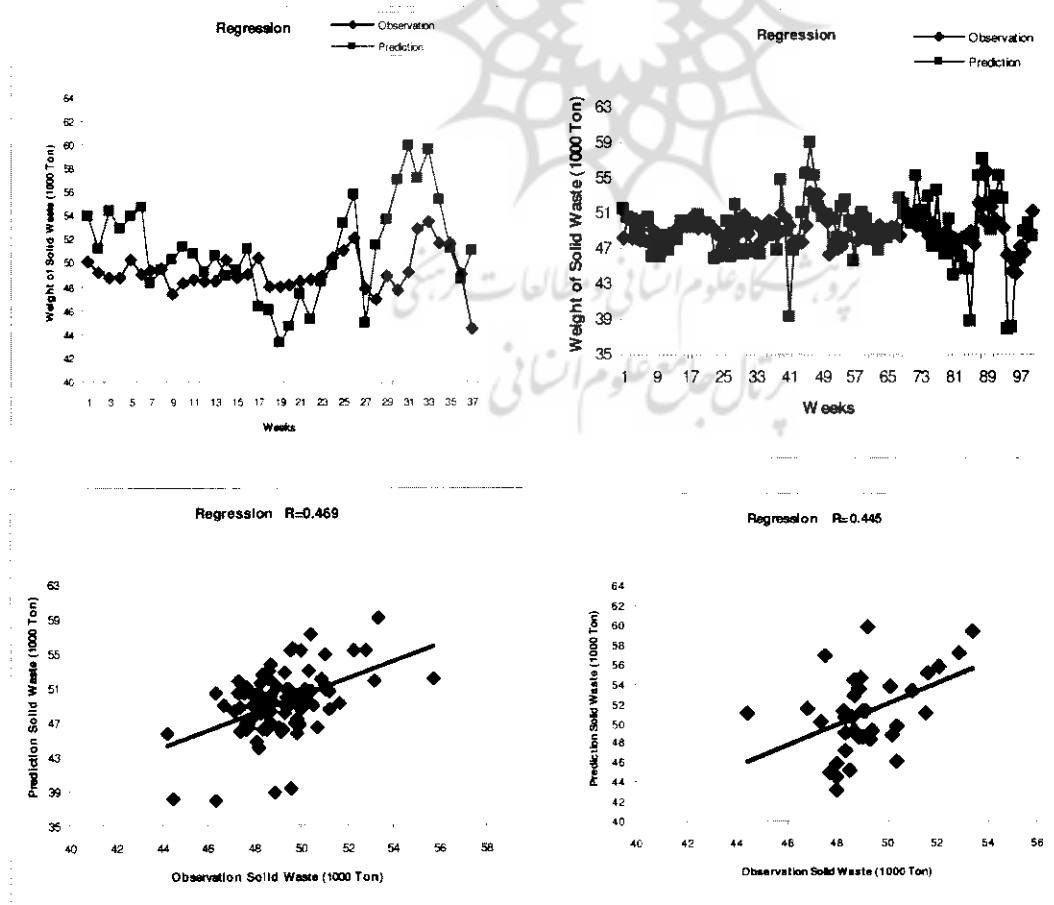
میکند. با توجه به جدول شماره بالا مشاهده می شود که مدل رگرسیون حاصل از داده های اصلی برای پارامترهای ورودی دارای مقادیر تورم واریانس و میزان تحمل نزدیک به یک یعنی مقدار ایده ال می باشد. در نهایت پس از پردازش اولیه بر روی داده های ورودی، مدل رگرسیونی با اجرای PCA برای تخمین مقدار زباله تولیدی تهیه گردید که معادله آن در زیر آورده شده است.

$$W = 49168332.29 + 1177698.532 * (PC3) - 904395.442 * (PC2) - 823336.065 * (PC9)$$

مطابق با معادله فوق پارامتر PC3 که در تشکیل آن وزن زباله سه هفته قبل ($W(t-3)$) بیشترین تاثیر را داشته است مهمترین عامل تاثیر گذار بر میزان زباله تولیدی شهر تهران می باشد. پارامترهای تاثیر گذار دیگر بر میزان زباله تولیدی به ترتیب عبارتند از: PC2 و PC9. همانطور که

همانطور که از جدول شماره بالا مشخص می باشد در مدل رگرسیونی با اجرای PCA از ۱۳ مولفه، تنها ورودی ۳ مولفه به مدل معنی دار بوده است و مدل مورد نظر مقدار زباله تولیدی را با توجه به این مقادیر ورودی برآورد

	R	MARE
train	0.469	0.048
test	0.445	0.066





میزان زباله تولیدی ساختارهای مختلفی از شبکه عصبی Feed-Forward با تعداد لایه‌های متفاوت و به ازای هر لایه با تعداد نرون های مختلف مورد بررسی قرار گرفت که در نهایت با توجه به معیار مورد بررسی R و MARE ساختار شامل سه لایه با تعداد نرون های ۱۳، ۲۲ و ۱ به ترتیب برای لایه اول، دوم و سوم به عنوان معماری برتر شبکه برگزیده شد. تابع آموزش و انتقال نیز به ترتیب TRAINLM و TANSIG مورد استفاده قرار گرفته شد. نتایج اجرای شبکه در مراحل train و test در جدول و شکل های زیر آورده شده است.

۳-۴- مقایسه دو مدل

همانطور که از جدول و شکل های بالا نیز مشخص است نتایج مطلوبی از مدل رگرسیونی به دست نیامده است و استفاده از مدل رگرسیونی برای تخمین میزان زباله تولیدی

	R	MARE
train	0.925	0.023846
test	0.837	0.043872

از جدول شماره () نیز مشخص است بیشترین تأثیر بر مولفه PC9 و PC2 را به ترتیب W(t-9) و W(t-12) داشته اند.

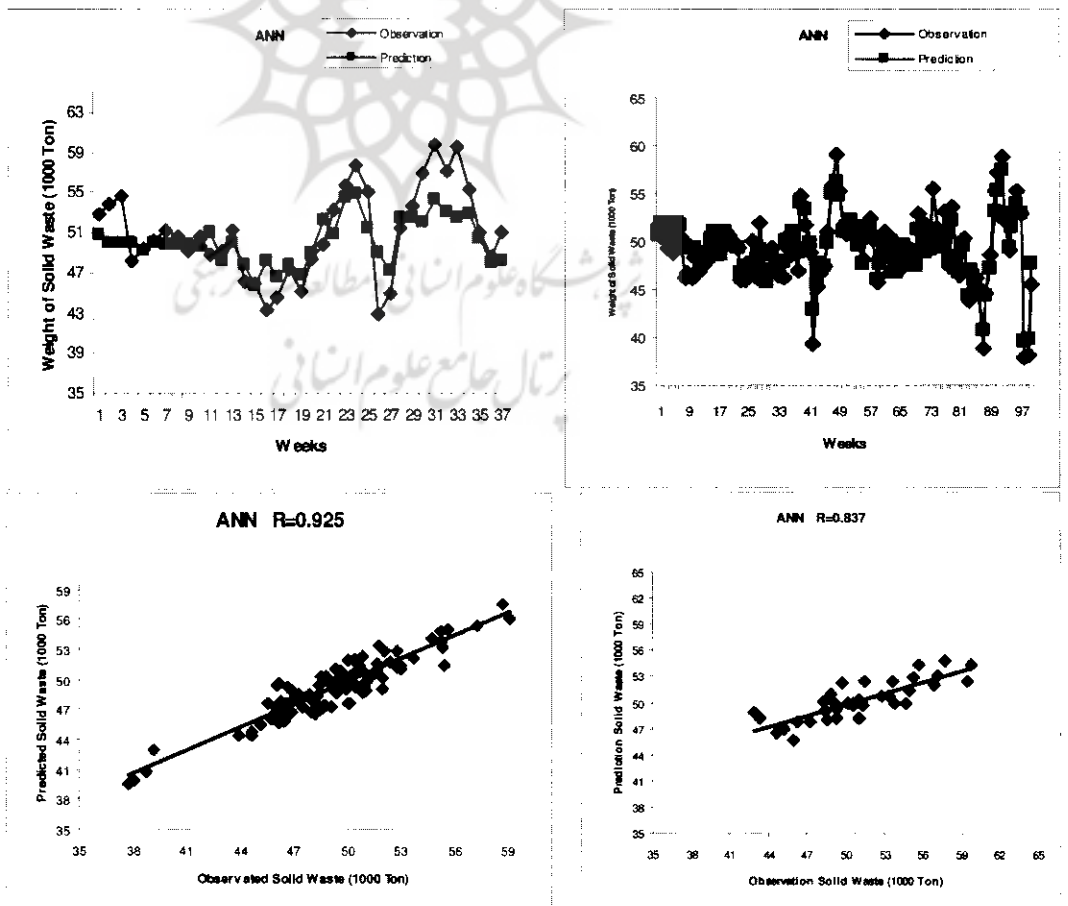
۳-۲-۱- ارزیابی اعتبار مدل رگرسیونی

پس از ساخت مدل رگرسیونی اعتبار مدل مورد بررسی قرار گرفت که نتایج مراحل train و test مدل در جدول و شکل های صفحه قبل آمده است.

همانطور که از جدول و شکل های صفحه قبل نیز مشخص است نتایج مطلوبی از مدل رگرسیونی به دست نیامده است و استفاده از مدل رگرسیونی برای تخمین میزان زباله تولیدی با خطای زیادی همراه است.

۳-۳- شبکه عصبی

برای رسیدن به بهترین ساختار شبکه جهت تخمین





منابع

1. Sahin, U., Ucan, O. N., Bayat, C., Oztorun, N., 2005. Modeling of SO₂ distribution in Istanbul using artificial neural networks. *Environmental Modeling and Assessment* 10:135-142.
2. Sahoo, G.B., Ray, C., De Carlo, E.H., 2006. Use of neural network to predict flash flood and attendant water qualities of a mountainous stream on Oahu, Hawaii. *Journal of Hydrology* 327: 525- 538
3. Shrestha, S., Kazama, F., 2007. Assessment of surface water quality using multivariate statistical techniques; A case study of the Fuji river basin, Japan. *Environmental Modelling & Software*, 22: 464e475.
4. Shrestha, S., Kazama, F., 2007. Assessment of surface water quality using multivariate statistical techniques; A case study of the Fuji river basin, Japan. *Environmental Modelling & Software*, 22: 464e475.
5. Karaca, F., Özkaya, B., 2006. NN-LEAP: A neural network-based model for controlling leachate flow-rate in a municipal solid waste landfill site. *Environmental Modelling & Software*, 21:1190-1197.
6. Dong, C., Jin, B., Li, D., 2003. Predicting the heating value of MSW with a feed forward neural network. *Waste Management*, 23: 103-106.
7. Lu, H.C., Hsieh, J.C., Chang, T.S., 2006. Prediction of daily maximum ozone concentrations from meteorological conditions using a two-stage neural network. *Atmospheric Research* 81: 124- 139.
8. Helena, B., Pardo, R., Vega, M., Barrado, E., Ferná' ndez, J.M., Ferná' ndez, L., 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. *Water Res.* 34, 807-816.
9. Hocking, R. R., *Methods and application of linear models regression and analysis of variance*, Wiley, Newjersey, 2003.
10. Legates, D.R. and McCabe, G.J. (1999). Evaluating the use of "Goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* 35, pp. 233-241.

با خطای زیادی همراه است ولی در عوض با استفاده از مدل شبکه عصبی نتایج بهبود زیادی نسبت به مدل رگرسیون یافته اند. به طور مثال مقادیر ضرایب همبستگی و میانگین نسبی خطای مطلق در مدل شبکه عصبی در حدود دو برابر بهتر از نتایج مشابه در مدل رگرسیونی است.

نکته دیگر این که در استفاده از مدل رگرسیونی و روش PCA فرض های زیادی وجود دارد که استفاده از آنها را جهت مسایل عملی مشکل می گرداند، مثلا همانطور که ذکر گردید متغیر وابسته در مدل رگرسیونی باید نرمال باشد و دیگر اینکه بین متغیرهای مستقل همبستگی زیادی وجود نداشته باشد. گذشته این مشکلات محدودیت مدل رگرسیونی جهت ورود متغیرها به مدل میباشد که در این مورد ورود متغیرهایی که معنی دار نیستند صورت نمی گیرد.

۴- نتیجه گیری

با توجه به سری زمانی زباله تولیدی برای شهر تهران که در شکل زیر مشاهده می شود هیچ روند خاصی در آن مشاهده نمی شود و نمودار نوسانات زیادی دارد با توجه به این موضوع احساس می شود که روش رگرسیون خطی چندمتغیره توانایی لحاظ کردن و آرایه مدلی که بتواند این نوسانات را لحاظ کند ندارد. و در این موارد استفاده از شبکه عصبی که قادر به پیش بینی روابط غیرخطی و پیچیده بین ورودیها و خروجی میباشد راه حل مناسبی جهت جایگزینی با رگرسیون خطی چندمتغیره می باشد.

رتال جامع علوم انسانی