## فصل اول: کلیات و تاریخچه



## فصل دوم: ممیزی بین دو مدل ARMA به همراه اغتشاش

## فصل سوم: توزیع آماری تابع ممیزی تقریبی در مدل ARMA به همراه اغتشاش



## فصل چهارم: معیارهای ممیزی در حوزه زمان

## فصل پنجم: موجک‌ها در ممیزی و خوشه‌بندی سری‌های زمانی

## فصل ششم: ممیزی داده‌های لرزه‌ای، نتیجه‌گیری و پیشنهادات

# فصل اول

# کلیات و تاریخچه

### ۱-۱) مقدمه

آنالیز ممیزی و خوشه‌بندی از مهمترین مباحث آماری هستند که کاربردهای وسیعی در علومی مانند مهندسی، زمین شناسی، پزشکی، اقتصاد، علوم رفتاری و غیره دارند. اولین بار در قرن هجدهم میلادی از آنالیز ممیزی برای دسته‌بندی گیاهان و حیوانات استفاده شد. مورانت[۱] (۱۹۲۶) مساله تخصیص یک جمجمه به عنوان کاسه سر یک اسکیمو یا یک نفر از نژاد انگلیسی را بررسی کرد. فیشر[۲] (۱۹۳۶) مساله ممیزی را فرمول بندی کرده و برای جدا سازی به کار گرفت. روش وی که امروزه نیز از روش های اساسی در علم ممیزی است بر مبنای تلاش برای یافتن یک معیار بهینه در جدا کردن دو جامعه آماری است. این معیار بهینه باید چنان باشد که تفاوت میانگین‌های دو جامعه را نسبت به واریانس مشترک آنها حداکثر کند. والد[۳] (۱۹۴٤) تأثیر جایگزینی پارامترهای نامعلوم با برآوردهای آنها بر توزیع تابع ممیزی را بصورتی جامع بررسی کرد.

قبل از فیشر، ممیزی نوعی آزمون فرض برای تساوی میانگین‌های دو یا چند توزیع تلقی می‌شد. از اینرو آماره آزمون به عنوان اندازه‌ای از تفاوت میان دو توزیع محسوب می‌شد و بر این پایه فاصله ماهالانوبیس به عنوان معیار واگرایی دو جامعه مطرح شد.

منظور از ممیزی ساختن قاعده‌ای است برای تخصیص یک عضو جدید به جامعه‌های از قبل معلوم و منظور از اصطلاح خوشه‌بندی، دسته‌بندی گروهی از اشیاء (متغیرها یا اقلام) است در حالیکه هیچ اطلاع اولیه‌ای از دسته‌های اشیاء وجود ندارد. کاربردهای ممیزی و خوشه‌بندی در علوم مختلف مانند اقتصاد، کشاورزی، گیاه شناسی، مهندسی، پزشکی، روانشناسی، پردازش تصویر[٤] و غیره چنان انبوه است که از حوصله این بحث خارج است. خواننده علاقه‌مند برای آشنای بیشتر با تئوری و کاربردها می‌تواند به یکی از کتاب‌های استاندارد در زمینه آنالیز چند متغیره پیوسته مانند جانسون و ویچرن[٥] (۱۹۹۲) مراجعه کند.

---
[۱] Morant
[۲] Fisher
[۳] Wald
[٤] Image Processing
[٥] Johnson and Wichern

با گسترش روز افزون علوم و نیازهای متفاوت در شاخه های مختلف دانش بشری همچون مهندسی و تکنولوژی شاخه های دیگری از علم در دههای اخیر به سرعت بنیان و گسترش یافته‌اند که در اهداف و عنوان بسیار به ممیزی و خوشه بندی نزدیک هستند مانند درخت‌های خوشه‌بندی[۱] یا تشخیص الگو[۲]. با وجود این نمی‌توان آنها را با ممیزی و خوشه‌بندی کاملاً یکسان دانست. در این رساله تعریف-های ارائه شده توسط هند[۳] (۱۹۸۱) که توسط اکثر محققین بعدی نیز پذیرفته شده‌اند، مد نظر قرار گرفته است.

### ۱-۲) تحلیل ممیزی و خوشه‌بندی در سری‌های زمانی

سری‌زمانی به مجموعه‌ای از مشاهدات گفته می‌شود که در طول زمان مرتب شده باشند. در سری زمانی، هر مشاهده با مشاهدات قبل وابستگی دارد. بررسی چنین مشاهداتی قسمت عظیمی از تجزیه و تحلیل و مدل‌بندی را در علم آمار به خود اختصاص داده است. جزئیات بسیاری در زمینه سری‌های زمانی منتشر گردیده و خواننده می‌تواند برای مطالعه جزئیات بیشتر به هنان[۴] (۱۹۶۰)، اندرسن[۵] (۱۹۷۱)، پریستلی[۶] (۱۹۸۱)، بروکول و دیویس[۷] (۱۹۹۱)، فولر[۸] (۱۹۹۶)، بریلینجر[۹] (۲۰۰۱) و چتفیلد[۱۰] (۲۰۰۳)،مراجعه کند. شاموی و استوفر[۱۱] (۲۰۱۱) نیز منبعی ارزشمند برای دیدن کاربردهای متنوع سری‌های زمانی می‌باشد.

یک سری زمانی غالباً یک الگو را تشکیل می‌دهد که می‌تواند پایه‌ای برای ممیزی یا خوشه‌بندی پدیده های مربوطه باشد. به عنوان مثال می‌توان مساله جدا سازی سیگنال از اغتشاش در مهندسی یا ممیزی

---

۱ Clustering Tree
۲ Pattern Recognition
۳ Hand
۴ Hannan
۵ Anderson
۶ Priestley
۷ Brockwell and Davis
۸ Fuller
۹ Brillinger
۱۰ Chatfield
۱۱ Shumway and Stoffer

داده‌های لرزه‌ای[۱] (داده‌های جمع آوری شده توسط لرزه نگار[۲] که در اثر زلزله یا انفجار تولید شده‌اند.) در زمین شناسی را نام برد. در ممیزی داده‌های لرزه‌ای هدف تخصیص یک سری مشاهده جدید لرزه-ای با منشاء ناشناس به زلزله یا انفجار می‌باشد. کاربردهای ممیزی سری‌های زمانی تنها محدود به علوم مهندسی یا زمین شناسی نیست . برای آشنایی بیشتر با برخی کاربردهای دیگر می‌توان به شاموی و استوفر (۲۰۱۱) مراجعه کرد.

شیوه های به کار گرفته شده در کتاب‌ها و مقالات برای ممیزی سری‌های زمانی را می‌توان در دو گروه مجزا جای داد.

گروه اول شامل روش‌های بهینه مرسوم در آمار یا مهندسی هستند، که معمولاً مبتنی بر پیش فرض نرمال بودن داده‌ها می‌باشند. در این روش‌ها متخصصین تلاش دارند یک معیار خوش تعریف از میزان خطای ممیزی را حداقل کنند.

گروه دوم شامل روش‌های مبتنی بر استخراج صورت‌های مناسب[۳] ممیزی است. منظور از صورت‌های مناسب کمیت‌هایی است که جوامع مختلف مورد بررسی از نظر آنها کاملاً یا تا حدود زیادی متفاوت هستند.

در دهه‌های اخیر خوشه‌بندی سری‌های زمانی به یکی از مهمترین عرصه‌های تحقیق در زمینه هایی همچون اقتصاد، بازاریابی، تجارت، مدیریت، پزشکی، فیزیک، روانشناسی و بسیاری دیگر از علوم تبدیل شده است. برای مثال در اقتصاد ممکن است به خوشه‌بندی کشورها بر پایه داده‌های اقتصادی همچون میزان رشد ناخالص ملی، میزان سرمایه گذاری خارجی، میزان بیکاری و یا نرخ تورم علاقه‌مند باشیم. به عنوان مثالی دیگر با جنبه‌ای متفاوت می‌توان به طبقه‌بندی بیماران قلبی بر پایه داده-های حاصل از دستگاه الکتروکاردیوگرافیک[٤] اشاره کرد. البته تمامی داده‌های ذکر شده در این دو مثال سری‌زمانی محسوب می‌شوند.

---

۱ Seismic Data
۲ Seismogram
۳ Feature Extracting
٤ Electro Cardio Graphics (ECG)

در مبحث خوشه‌بندی سری‌های زمانی تمرکز مطالعات انجام شده بر یافتن کمیت‌های مناسب از داده ها برای اندازه‌گیری مشابهت‌ها و تفاوت‌ها در داده‌های سری‌زمانی است. سپس با استفاده از فاصله اقلیدسی، معیار واگرایی کولبک – لیبلر[۱]، معیار واگرایی چرنوف[۲] یا دیگر معیارهای متقارن از فواصل، داده‌ها را به چند گروه تقسیم می‌کنند.

### ۱–۳) توصیف عام مساله ممیزی

مساله تخصیص بردار سری‌زمانی $\mathbf{x}' = (x_1,\ldots,x_T)$ را در حالت کلی می‌توان به صورت زیر تشریح نمود. سری زمانی مشاهده شده $\mathbf{x}$، به یکی از $g$ جامعه $\Pi_1,\Pi_2,\ldots,\Pi_g$ تعلق دارد. حال مساله این است که به طریقی بهینه، $\mathbf{x}$ به یکی از این جوامع تخصیص داده شود.

فرض کنید سری مشاهده شده $\mathbf{x}$ را داریم و می‌خواهیم بررسی کنیم که به کدامیک از $g$ جامعه $\Pi_i$ با چگالی متناظر $p_i(\mathbf{x})$ برای $i=1,\ldots,g$، متعلق است. فضایی که سری $\mathbf{x}$ می‌تواند در آن مشاهده شود، به صورت $R_1,R_2,\ldots,R_g$ افراز می‌شود. $\mathbf{x}$ را به جامعه $\Pi_i$ تخصیص می‌دهیم اگر در ناحیه $R_i$ قرار گیرد. احتمال تخصیص اشتباه عبارت است از احتمال تخصیص سری $\mathbf{x}$ به جامعه $\Pi_j$ زمانی که در واقع به جامعه $\Pi_i$ تعلق دارد و به صورت زیر محاسبه می‌شود (برای $i \neq j$)

$$P(j|i) = \int_{R_j} p_i(\mathbf{x})\, d\mathbf{x} \qquad (۱–۱)$$

فرض کنید احتمال پیشین اینکه سری $\mathbf{x}$ متعلق به جامعه $\Pi_i$ باشد، برای $i=1,\ldots,g$ برابر $\pi_i$ باشد. در این صورت احتمال آنکه سری $\mathbf{x}$ متعلق به $\Pi_i$ باشد و به اشتباه به $\Pi_j$ تخصیص داده شود برابر با $\pi_i P(j|i)$ است و کل احتمال تخصیص اشتباه عبارت است از

$$P_e = \sum_{i=1}^{g} \pi_i \sum_{j\neq i} P(j|i) \qquad (۲–۱)$$

---
[۱] Kullback - Libler
[۲] Chernoff

می‌توان هزینه را نیز در فرمول (۱–۲) جای داد. برای این منظور کافی است $P(j|\,i)$ را در $C(j|i)$ (هزینه ناشی از تشخیص یک سری از جامعه $\Pi_i$ به عنوان عضوی از جامعه $\Pi_j$) ضرب کرد.

اگر هزینه ناشی از تخصیص اشتباه در دو جامعه را یکسان بگیریم، آنگاه احتمال کل تخصیص اشتباه با تخصیص $\mathbf{x}$ به جامعه $\Pi_i$ حداقل می شود، هرگاه برای تمامی $i \neq j$ ها داشته باشیم (اندرسن، ۱۹۸٤)

$$\frac{p_i(\mathbf{x})}{p_j(\mathbf{x})} > \frac{\pi_j}{\pi_i}. \qquad (۱–۳)$$

در حالت کلی تر برای $g$ جامعه، $\mathbf{x}$ را به جامعه $\Pi_i$ی تخصیص می‌دهیم اگر بزرگترین احتمال پسین یعنی،

$$P(\Pi_i|\,\mathbf{x}) = \frac{\pi_i\; p_i(\mathbf{x})}{\sum_j \pi_j(\mathbf{x}) p_j(\mathbf{x})}, \qquad (۱–٤)$$

را داشته باشد. می‌توان گفت که قاعده تخصیص $\mathbf{x}$ بر اساس (۱–۳) یا (۱–٤) معادل هستند.

با روش کلاسیک نیمن – پیرسن برای دو جامعه، $\mathbf{x}$ به جامعه اول تخصیص داده می‌شود اگر

$$\frac{p_۱(\mathbf{x})}{p_۲(\mathbf{x})} > K, \qquad (۱–۵)$$

که در آن $K$ با توجه به میزان احتمال خطای نوع اول تعیین می‌شود. این قاعده معادل با قاعده بیز است اگر $K = \frac{\pi_۲}{\pi_۱}$.

اگر فرض کنیم تحت $\Pi_j$ ($j = ۱,\ldots,g$) بردار $\mathbf{x}$ دارای توزیع نرمال $p$ متغیره با بردار میانگین $\boldsymbol{\mu}_j$ و ماتریس کوواریانس $\boldsymbol{\Sigma}_j$ است، آنگاه

$$p_j(\mathbf{x}) = (۲\pi)^{-\frac{p}{۲}} \left|\boldsymbol{\Sigma}_j\right|^{-\frac{۱}{۲}} \exp\left\{-\frac{۱}{۲}(\mathbf{x}-\boldsymbol{\mu}_j)'\boldsymbol{\Sigma}_j^{-۱}(\mathbf{x}-\boldsymbol{\mu}_j)\right\}, \qquad \mathbf{x} \in R^p \qquad (۱–٦)$$

برای ساختن قاعده ممیزی از کمیتی متناسب با لگاریتم تابع درستنمایی به صورت زیر استفاده می شود،

$$\ell_j(x) = -\frac{۱}{۲}\ln\left|\boldsymbol{\Sigma}_j\right| - \frac{۱}{۲}\mathbf{x}'\boldsymbol{\Sigma}_j^{-۱}\mathbf{x} + \boldsymbol{\mu}'_j\boldsymbol{\Sigma}_j^{-۱}\mathbf{x} - \frac{۱}{۲}\boldsymbol{\mu}'_j\boldsymbol{\Sigma}_j^{-۱}\boldsymbol{\mu}_j - \frac{p}{۲}\ln(۲\pi_j) \qquad (۱–۷)$$

معمولاً در عباراتی که شامل لگاریتم درستنمایی است از عبارت ثابت $-\ln(۲\pi)$ صرف نظر می‌شود. در این مورد $\mathbf{x}$ به جامعه $\Pi_i$ تعلق می‌گیرد اگر برای $i \neq j$ ، $i = ۱,\ldots,g$ داشته باشیم:

$$\ell_i(\mathbf{x}) > \ell_j(\mathbf{x}) \qquad (۱-۸)$$

با توجه به اینکه نتایج برای $g = ۲$ تفاوتی با بیش از دو جامعه ندارد فرض می‌شود $g = ۲$. در حالتی که ماتریس‌های کواریانس دوجامعه یکسان باشند یعنی، $\mathbf{\Sigma}_۱ = \mathbf{\Sigma}_۲ = \mathbf{\Sigma}$. ، می‌توان ضابطه (۱-۸) را برحسب تابع ممیزی خطی زیر بیان کرد

$$\begin{aligned} d_l(\mathbf{x}) &= \ell_۱(\mathbf{x}) - \ell_۲(\mathbf{x}) \\ &= (\boldsymbol{\mu}_۱ - \boldsymbol{\mu}_۲)'\mathbf{\Sigma}^{-۱}\mathbf{x} - \frac{۱}{۲}(\boldsymbol{\mu}_۱ - \boldsymbol{\mu}_۲)'\mathbf{\Sigma}^{-۱}(\boldsymbol{\mu}_۱ - \boldsymbol{\mu}_۲) + \ln\frac{\pi_۱}{\pi_۲}, \end{aligned} \qquad (۱-۹)$$

که در آن $\mathbf{x}$ به جامعه $\Pi_۱$ یا $\Pi_۲$ تعلق می‌گیرد اگر به ترتیب $d_l(\mathbf{x}) \geq ۰$ یا $d_l(\mathbf{x}) < ۰$. تابع ممیزی خطی ترکیبی از متغیرهای با توزیع نرمال است و به سادگی می‌توان نشان داد که با فرض تعلق داشتن مشاهدات به جامعه اول دارای میانگین $\frac{D^۲}{۲}$ و واریانس $D^۲$ و با فرض اینکه مشاهدات متعلق به جامعه دوم هستند دارای میانگین $-\frac{D^۲}{۲}$ و واریانس $D^۲$ است که در آن $D^۲$ فاصله ماهالانوبیس بین $\boldsymbol{\mu}_۱$ و $\boldsymbol{\mu}_۲$ است و به صورت زیر تعریف می‌شود

$$D^۲ = (\boldsymbol{\mu}_۱ - \boldsymbol{\mu}_۲)'\mathbf{\Sigma}^{-۱}(\boldsymbol{\mu}_۱ - \boldsymbol{\mu}_۲) \qquad (۱-۱۰)$$

بنابراین با توجه به رابطه (۱-۱) احتمال تخصیص اشتباه عبارت است از

$$P(۱|۲) = P(۲|۱) = \Phi\left(-\frac{D}{۲}\right) \qquad (۱-۱۱)$$

که در آن $\Phi(z)$ معرف توزیع تجمعی نرمال استاندارد است. می‌توان نتیجه گرفت که کارایی تابع ممیزی خطی وابسته به فاصله ماهالانوبیس بین میانگین دو جامعه است.

در حالتی که ماتریس کواریانس دو جامعه متفاوت باشد تابع ممیزی رابطه (۱-۹) منجر به یک تابع درجه دوم به شکل زیر می‌شود،

$$d_Q(\mathbf{x}) = -\frac{1}{2}\ln\frac{|\mathbf{\Sigma}_1|}{|\mathbf{\Sigma}_2|} - \frac{1}{2}\mathbf{x}'(\mathbf{\Sigma}_1^{-1} - \mathbf{\Sigma}_2^{-1})\mathbf{x} + (\boldsymbol{\mu}_1'\mathbf{\Sigma}_1^{-1} - \boldsymbol{\mu}_2'\mathbf{\Sigma}_2^{-1})\mathbf{x} + \ln\frac{\pi_1}{\pi_2} \qquad (۱۲-۱)$$

بر خلاف تابع ممیزی خطی، تابع ممیزی درجه دو دارای توزیع بسیار پیچیده‌ای است. در بخش بعد به این موضوع بیشتر پرداخته خواهد شد.

در حالتی که پارامترهای $\boldsymbol{\mu}_i$ و $\mathbf{\Sigma}_i$ نامعلوم هستند می‌توان با گرفتن یک نمونه تصادفی مانند $\mathbf{x}_{i1},\ldots,\mathbf{x}_{iN_i}$ به حجم $N_i$ از جامعه $\Pi_i$، بردار میانگین و ماتریس کواریانس را برای هر یک از گروه های $i=1,2,\ldots,g$ برآورد کرد. در این صورت برآورد $\boldsymbol{\mu}_i$، $\overline{\mathbf{x}}_i = \frac{1}{N_i}\sum_{j=1}^{N_i}\mathbf{x}_{ij}$ و برآورد $\mathbf{\Sigma}_i$،

$$S_i = (N_i - 1)^{-1}\sum_{j=1}^{N_i}(\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)(\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)' \qquad (۱۳-۱)$$

می باشد (اندرسن، ۱۹۸٤). هنگامی که فرض می‌شود ماتریس کواریانس جامعه یکسان هستند، برآورد $\mathbf{\Sigma}$ به صورت زیر خواهد بود

$$S = \left(\sum_i N_i - g\right)^{-1}\sum_i (N_i - 1)\, S_i \qquad (۱٤-۱)$$

برای کاربرد در تابع ممیزی خطی می‌توان با استفاده از برآوردهای بالا، $g_i(\mathbf{x})$ در رابطه (۱-۷) را به صورت زیر برآورد کرد

$$\hat{g}_i(\mathbf{x}) = \overline{\mathbf{x}}_i'\, S^{-1}\,\mathbf{x} - \frac{1}{2}\overline{\mathbf{x}}_i'\, S^{-1}\,\overline{\mathbf{x}}_i + \ln\pi_i \qquad (۱٥-۱)$$

از آنجا که $\overline{\mathbf{x}}_i$ و $S$ در احتمال به $\boldsymbol{\mu}_i$ و $\Sigma_i$ همگرا هستند بنابراین $\hat{g}_i(\mathbf{x})$ نیز در احتمال به $g_i(\mathbf{x})$ همگرا است (شاموی و استوفر، ۲۰۱۱).

عملکرد تابع ممیزی خطی را می‌توان به طرق مختلف ارزیابی کرد. اگر پارامترهای جامعه معلوم باشند، می‌توان با استفاده از (۱-۱۰) و (۱-۱۱) احتمال خطای ممیزی را محاسبه کرد. اگر پارامترها مجهول باشند می‌توان از برآورد فاصله ماهالانوبیس در روابط (۱-۱۰) یا (۱-۱۱) استفاده کرد به شرط آنکه

نمونه به اندازه کافی بزرگ باشد. یک راه دیگر برای ارزیابی عملکرد تابع ممیزی از محاسبه نرخ خطا[1] در تخصیص نمونه‌ها توسط قاعده ممیزی است. اگر $n_{ij}$ مشخص کننده تعداد مشاهداتی از جامعه $\Pi_j$ باشد که به اشتباه به جامعه $\Pi_i$ تخصیص داده شده اند آنگاه نرخ خطای نمونه‌ای را می‌توان برای $i \neq j$ با استفاده از نسبت زیر به‌دست آورد.

$$\hat{P}(i \mid j) = \frac{n_{ij}}{\sum_i n_{ij}} \qquad (۱-۱٦)$$

با توجه به اینکه درمحاسبه نرخ خطا از همان نمونه‌هایی استفاده شده است که در ساختن قاعده ممیزی به کار گرفته شده‌اند لذا برای رفع این عیب و به‌دست آوردن برآوردی واقعی تر از میزان خطای ممیزی می‌توان از روش پیشنهادی لاچنبروچ و میکی[2] (۱۹٦۸) استفاده کرد که در آن ابتدا یکی از نمونه‌ها را کنار گذاشته و تابع ممیزی را بر اساس نمونه‌های باقی مانده محاسبه کرده سپس نمونه کنار گذاشته شده را با تابع ممیزی تخصیص می‌دهند. این روند برای تمامی نمونه‌ها تکرار می‌شود و با استفاده از رابطه (۱-۱٦) نرخ خطای ممیزی محاسبه می‌شود. این روش را روش یکی را کنار گذاشتن[3] می‌گویند.

### ۱-٤) تاریخچه

مانند دیگر مباحث سری‌های زمانی، آنالیز ممیزی سری‌زمانی نیز در دو بخش حوزه زمان و حوزه فرکانس مطرح است که در هر دو حوزه مطالعات وسیعی انجام گرفته است. در حوزهٔ زمان نسبت لگاریتم درستنمایی توابع چگالی دو جامعه به عنوان معیاری برای ممیزی بین دو مدل مورد استفاده قرار می‌گیرد. هر گاه تفاوت دو جامعه تنها در میانگین‌های دو جامعه باشد ممیزی به یک رابطه خطی منجر می‌شود. ممیزی خطی در کتاب‌های مختلف کلاسیک مانند به اندرسن (۱۹۸٤) مورد بررسی قرار گرفته است. اما اگر تفاوت بین واریانس‌های دو جامعه باشد ممیزی منجر به یک تابع درجه دوم می‌شود که دارای توزیع بسیار پیچیده‌ای است. می‌توان نشان داد که این تابع ترکیب خطی از متغیرهای

[1] Error Rate
[2] Lachenbruch and Mickey
[3] Leave One Out

تصادفی کای‌دو هر کدام با یک درجه آزادی است (چان[1] و همکاران، ۱۹۹٦). اما به‌دست آوردن ضرایب ترکیب خطی نیاز به اعمالی مانند حاصل‌ضرب ماتریس‌ها و عکس ماتریس‌ها دارد. در این حالت تابع ممیزی به صورت تحلیلی به‌دست نمی‌آید و باید با روش عددی و تقریبی محاسبه گردد. چون ماتریس‌های کواریانس سری‌های زمانی دارای ابعاد بزرگ هستند روش‌های عددی نیز راه حل خوبی برای مساله نیستند. در حالت خاص ممیزی بین دو سری‌زمانی خودبازگشتی (AR) چان (۱۹۹۱) و خودبازگشت میانگین متحرک (ARMA) چان و همکاران (۱۹۹٦) تابع ممیزی تحلیلی تقریبی را به‌دست آورده‌اند. آنها در ضمن نشان داده‌اند که در ممیزی بین دو مدل میانگین متحرک (MA) با مرتبه یک می‌توان تابع ممیزی دقیقی را به‌دست آورد. توسعه این کار توسط چینی پرداز (۲۰۰۰) در اتورگرسیو مرتبه اول همراه با یک اغتشاش و به وسیله چینی پرداز و کاکس (۲۰۰٤) نشان داده شده است.

در حالتی که پارامترهای جامعه معلوم هستند، می‌توان نشان داد که ممیزی خطی بهینه است؛ اما اگر پارمترها برآورد شوند، خصوصاً زمانی که ابعاد جامعه نسبت به حجم نمونه بزرگ می‌باشد ممیزی خطی بهینه نخواهد بود (شاموی و استوفر، ۲۰۱۱). در ممیزی درجه دوم نیز همان‌گونه که قبلاً ذکر شد کار کردن در حوزه زمان به واسطه برخورد با ماتریس‌های با ابعاد بسیار بزرگ مشکل می‌باشد. به این دلیل مطالعات در زمینه ممیزی سری‌های زمانی بر حوزه طیفی سری‌های زمانی متمرکز بوده است. در این حوزه آنالیز ممیزی با استفاده از تقریب های طیفی دنبال می‌شود. لیگت[2] (۱۹۷۱) برای به‌دست آوردن تابع ممیزی از روش طیفی استفاده نمود. وی حالتی را در نظر گرفت که میانگین‌های دو جامعه برابر باشند. سپس با استفاده از تقریب‌های طیفی به ممیزی بین سری‌های زمانی پرداخت. شاموی و آنگر[3] (۱۹۷٤) در ادامهٔ کار اندرسن (۱۹۷۱)، کولبک (۱۹۷۸) و اندرسن و بهادر[4] (۱۹٦۲) با استفاده از معیار اطلاع ممیزی کولبک – لیبلر، بهترین توابع ممیزی خطی را برای ممیزی بین دو فرآیند ایستای

[1] Chan
[2] Liget
[3] Shumway and Unger
[4] Anderson and Bahudur

نرمال، زمانی که میانگین‌های دو فرایند برابر هستند، فراهم نمودند. آنها نتایج این روش‌ها را برای ممیزی بین داده‌های لرزه‌ای (داده‌های ثبت شده توسط لرزه نگار که بر اثر نوسانات ناشی از زمین لرزه یا انفجار حاصل شده‌اند) ارائه کردند. از دیگر افرادی که در زمینه آنالیز ممیزی در حوزه طیفی سری-های زمانی مطالعه کردند می‌توان به آلاگان[۱] (۱۹۸٦) درگاهی – نوبری[۲] (۱۹۹۲) درگاهی – نوبری و لی‌کاک[۳] (۱۹۸۱) ، کاکیزاوا و همکاران[٤] (۱۹۹۸) اشاره کرد. در بیشتر این مقالات تابع ممیزی با استفاده از روش نسبت درستنمایی و یا ماکزیمم کردن فاصله‌هایی مانند کولبک–لیبلر، باتاچاریا و چرنوف و با فرض ایستایی به دست آمده است. در عمل بسیاری از سری‌های مشاهده شده را نمی‌توان به‌صورت ایستا مدل‌بندی کرد. به عنوان مثالی از سری‌های غیر ایستا می‌توان به داده‌های لرزه‌ای اشاره کرد.

ایده اساسی در مدل بندی سری‌های غیر ایستا بر گرفته از نمایش کرامر– فوریه[٥] یک فرآیند ایستا می-باشد. هر فرآیند ایستای مرتبه دوم مانند $x_t$ با میانگین صفر را می توان به صورت $x_t = \int_{(-\pi,\pi]} A(\omega)\exp(i\omega t)\,dZ(\omega), \quad t \in Z$ نمایش داد، که در آن $A(\omega)$ تابع انتقال[٦] و $Z(\omega)$ یک فرآیند تصادفی با میانگین صفر و نمو[۷] متعامد است. پریستلی (۱۹٦٥) یک نمایش معادل را برای فرآیندهای غیر ایستا ارائه کرد که در آن تابع انتقال $A(\omega)$ تابعی از زمان است. دهلوس[۸] (۱۹۹۷) با ایده گرفتن از این راهبرد، مدل‌بندی ایستای موضعی[۹] را مطرح کرد که به وسیله آن تابع انتقال وابسته به زمان را بر یک فاصله فشرده تعریف نمود. اومباو[۱۰] و همکاران (۲۰۰۱ ، ۲۰۰۲) روش هموارسازی نمایی مرکب موضعی[۱۱](SLEX) را پیشنهاد نمودند که در آن یک سری‌زمانی ایستای موضعی به قسمت‌های دو به دو جدا افراز می‌شود به طوری که هر قسمت ایستا است و هر قسمت را می‌توان

---
۱ Alagon
۲ Dargahi - Noubary
۳ Dargahi – Noubary and Laycock
٤ Kakizawa et. al.
٥ Cramer – Fourier Representation
٦ Transfer Function
۷ Increment
۸ Dahlhaus
۹ Locally Stationary
۱۰ Ombao
۱۱ Smooth Localized Complex Exponential

مانند نمایش کرامر- فوریه یک سری ایستا مدل‌بندی کرد با این تفاوت عمده که عبارات فوریه $\exp(i\omega t)$ با عبارت تعدیل شده متناسب با هر قسمت عوض می‌شود.

ایده های مطرح شده برای مدل بندی سری‌های زمانی ایستای موضعی را می‌توان در ممیزی و خوشه بندی سری‌های زمانی غیر ایستا به خدمت گرفت. بر این اساس ساکایاما و تانی گوچی[۱] (۲۰۰٤) در حالت چند متغیره و با فرض نرمال بودن سری یک روش ممیزی را برای فرآیندهای ایستای موضعی دهلوس (۱۹۹۷) بر اساس معیار واگرایی کولبک – لیبلر ارائه نمودند. شاموی (۲۰۰۳) بعضی از وجوه این روند را برای حالت یک متغیره بررسی نمود. هوانگ و همکاران[۲] (۲۰۰٤) یک روش ممیزی را برای فرآیندهای ایستای موضعی در مدل‌های SLEX معرفی نمودند. روش آنها مشتمل بر انتخاب یک پایه مناسب جهت ممیزی بین دو جامعه و سپس ساختن تابع ممیزی بر اساس معیار کولبک – لیبلر، در پایه انتخابی است. چاندلر و پولونیک[۳] (۲۰۰٦) برای یک فرآیند اتو رگرسیو بطور موضعی ایستا ( با واریانس متغیر در زمان) یک قاعده ممیزی پیشنهاد کردند که نه بر پایه اندازه‌ای از فاصله بلکه بر پایه شکل تابع واریانس می‌باشد.

موجک‌ها[٤] به دلیل تمرکز همزمان بر فرکانس و زمان، ابزار جایگزین مناسبی برای پایه‌های فوریه در مدل‌بندی پدیده‌های غیر ایستا می‌باشند. پدیده‌هایی که تابع طیف آنها در زمان تغییر می‌کند. به عنوان یک منبع مناسب جهت آشنایی با موجک‌ها و کاربرد آنها در آمار می‌توان نیسن[٥] (۲۰۰۸) را دید. نیسن و همکاران (۲۰۰۰) موجک ایستای موضعی[٦] (LSW) را برای مدل‌بندی سری‌های زمانی معرفی کردند. مدل LSW فرآیند تحت بررسی را در زمان و مقیاس[۷] تجزیه می‌کند و امکان برآورد تابع طیف موجکی تکاملی[۸] (EWS) را فراهم می‌کند. فریزلویکس و اومباو[۹] (۲۰۰۹) با استفاده از EWS

---

[۱] Sakiyama and Taniguchi
[۲] Huang et.al.
[۳] Chandler and Polonik
[٤] Wavelet
[٥] Nason
[٦] Locally Stationary Wavelet
[۷] Scale
[۸] Evolutionary Wavelet Spectrum
[۹] Fryzlewicz and Ombao

به ممیزی در سری‌های زمانی پرداختند. به دلیل خواص موجک‌ها، برآورد کمیت‌هایی مانند تابع طیف موجکی بسیار بهینه‌تر از برآوردهای نظیر آنها مانند تابع طیف موضعی روش دهلوس می‌باشند. همچنین مدل SLEX دارای این محدودیت است که می‌باید فرایند تحت بررسی به یک مضرب دو افراز شود در حالی که مدل LSW دارای این محدودیت نیست. چینی پرداز (۲۰۰۱) ممیزی یک سری زمانی گوسی را در فضای حالت[۱] انجام داد. از دیگر مقالات ارائه شده در زمینه ممیزی سری های زمانی می‌توان به آلونسو و همکاران (۲۰۰۶) اشاره کرد. همچنین مهاراجه و آلنسو[۲] (۲۰۰۷) برای ممیزی در فرآیندهای ایستای موضعی از واریانس موجک‌ها استفاده کردند.

موجک‌ها و بسته‌های موجک[۳] را علاوه بر ممیزی در خوشه‌بندی نیز می‌توان به کار گرفت. می‌یر و چینروگ رونگ[٤] (۲۰۰۳) با استفاده از معیار کولبک- لیبلر روشی را برای خوشه بندی سیگنال‌ها بر اساس ضرایب موجکی پیشنهاد کردند. در یک کار مرتبط ونوسی و همکاران[٥] (۲۰۰۵) یک روش بیزی برای انتخاب بسته های موجک مناسب برای خوشه‌بندی پیشنهاد کردند.

### ۱-۵) معرفی فصل‌ها

در فصل دوم این رساله کارهای چان و همکاران (۱۹۹٦) و چینی پرداز (۲۰۰۰) در مدل‌های AR(p) بسط داده شده است. به این ترتیب که ابتدا ممیزی بین دو مدل AR(p) به همراه اغتشاش برای حالتی که در آن واریانس‌های دو مدل و دو اغتشاش یکسان نیستند بررسی شده و سپس تابع ممیزی بین دو مدل MA(۱) به همراه اغتشاش با واریانس های متفاوت به دست آمده است (منصوری و چینی پرداز، ۲۰۱۰). در این فصل عملکرد تابع ممیزی برای مقادیر مختلف پارامتر و واریانس با استفاده از شبیه سازی در مدل های ARMA نشان داده شده است. همچنین ممیزی مدل‌های آمیخته بررسی شده و در مدل‌های اتورگرسیو بسط نتایج به حالت غیر ایستا نشان داده شده است.

---

۱ State Space
۲ Alonso and Maharaj
۳ Wavelet Packets
٤ Meyer and Chinrungrueng
٥ Vannucci et al.

در فصل سوم توزیع تابع ممیزی یا کومولانت‌های تابع ممیزی برای مدل های ARMA به همراه اغتشاش به‌دست آمده و کومولانت‌های تابع ممیزی با آنچه دیگران به صورت عددی به دست آورده اند مقایسه شده است.

در فصل چهارم آنالیز ممیزی سری‌های زمانی با استفاده از معیارهای جداسازی مانند کولبک- لیبلر و چرنوف مورد بررسی قرار گرفته است. به این ترتیب که در ابتدا معیارهای ممیزی در حوزه فرکانس بررسی شده است (چینی پرداز و همکاران، ۱۳۸۸). سپس معیارهای ممیزی در حوزه زمان به‌دست آمده و عملکرد تقریب‌های به‌دست آمده برای معیارهای کولبک- لیبلر و چرنوف در حوزه زمان با تقریب‌های مشابه در حوزه فرکانس با استفاده از شبیه‌سازی مقایسه شده است. نتایج به‌دست آمده نشان از برتری تقریب‌ها در حوزه زمان دارد.

در فصل پنجم موجک‌ها به اختصار معرفی شده و کاربرد آنها در خوشه‌بندی و ممیزی سری‌های زمانی نشان داده شده است. یک روش برای خوشه‌بندی سری‌های زمانی پیشنهاد شده و نتایج با کار دیگران مقایسه شده است. همچنین برای یک سری‌زمانی ایستای نرمال تابع درستنمایی بر اساس ضرایب موجک گسسته به‌دست آمده و با استفاده از آن یک معیار جدید جهت ممیزی سری‌های زمانی با استفاده از موجک ها معرفی شده و عملکرد آن در ممیزی سری‌های زمانی با استفاده از شبیه‌سازی نشان داده شده است.

در فصل آخر این رساله یک روش جهت ممیزی و خوشه‌بندی داده‌های لرزه‌ای با استفاده از موجک ها معرفی شده است و در ادامه به بحث و نتیجه گیری پیرامون روش‌های به کار رفته، پرداخته شده و پیشنهاداتی برای کارهای آینده ارائه شده است.