سیزدهمین کنفرانس
آمار ایران
۲-۴ شهریور ۱۳۹۵

خوشه‌بندی زوایای تصادفی با مدل آمیخته توزیع فون‌میزس

سیما نوری جویباری [۱، ۲]، آناهیتا نودهی [۲] و موسی گل‌علی‌زاده [۳]

کارشناس ارشد آمار [۱]

دانشجوی دکتری آمار، گروه آمار، دانشگاه تربیت مدرس [۲]

عضو هیئت علمی، گروه آمار، دانشگاه تربیت مدرس [۳]

چکیده در تحقیقات علوم زیستی نشان داده شد که زوج زاویه‌ای وجود دارد که تا حد دقیقی ساختار هندسی و فضایی کامل یک پروتئین را در یک فضای سه بعدی توصیف می‌کنند. برای تشریح احتمالاتی بر اساس موقعیت اتمهای پروتئین و فضای قرارگیری زوایا نشان داده شد که چنبره است، مقادیر توام این دو زاویه دارای توزیعی به نام فون‌میزس دومتغیره هستند. با درنظر گرفتن یک مدل آمیخته از این توزیع می توان به مدل بندی احتمالی مجموعه ای از زوایای با وزن های متناسب پرداخت. به عبارتی دیگر، با این رویکرد امکان مدل‌بندی توزیع‌های آمیخته در حالت مدهای چندگانه نیز وجود خواهد داشت. این موضوع زمینه مناسبی برای خوشه بندی مجموعه ای از پروتئین ها فراهم می کند. لذا، هدف اصلی مقاله حاضر بررسی خوشه‌بندی زوایای تصادفی بر اساس مدل آمیخته توزیع فون میزس دومتغیره، نحوه برآورد پارامترهای آن‌ها و کاربست این مدل برای یک مثال واقعی با روش‌های است.

واژه‌های کلیدی: زوایای تصادفی، توزیع فون میزس دو متغیره، مدل آمیخته، خوشه‌بندی.

کد موضوع‌بندی ریاضی (۲۰۱۰): 62H11، 62H30.

---

۱ مقدمه

اگرچه آمار و احتمال نقش بی بدیلی در پیشرفت نظریه‌های کاربردی و نظری بر اساس متغیرهای تصادفی در فضای اقلیدسی ایفا کردند، اما پیشرفت تکنولوژی، پدیده‌های تصادفی جدیدی را پیش روی بشر قرار داد که توزیع‌های احتمالاتی (اقلیدسی) قادر به مدل‌بندی آن‌ها نیستند. به عنوان مثال، در طی دهه‌های گذشته استفاده از زوایا توانسته جای پای محکمی در تحلیل آماری داده‌هایی که فضای رخداد آنها نااقلیدسی است، باز کند. نظریه آماری که به اینگونه مسائل می‌پردازد به آمار غیرخطی معروف است.

از نظر مطالعات مروری، پایه‌ریزی تئوری‌های احتمالاتی مربوط به زوج زوایا که مقادیرشان را در بازه $[0, 2\pi)$ اختیار می‌کنند، اولین

---

[۲] نام ارائه دهنده مقاله: سیما نوری جویباری، s.nouri@modares.ac.ir

بار توسط ماردیا (۱۹۷۵) با معرفی توزیع فون میزس دو متغیره[۱] با هشت پارامتر شروع شد. لازم به اشاره است مجموعه نقاط هندسی حاصل از تغییرات توام این دو زاویه در بازه $[0, 2\pi]$ تشکیل چنبره می دهد. علاوه براین، در محافل زیستی این زوج زاویه به زوایای دوسطحی معروفند (راماچاندران و همکاران (۱۹۶۳)). هدف اصلی معرفی توزیعی شبیه توزیع نرمال دومتغیره بود. سال‌ها بعد، حالت‌های خاصی از توزیع فون میزس دو متغیره با عنوان مدل سینوسی با پنج پارامتر توسط سینگ و همکاران (۲۰۰۲) معرفی شد. پس از آن با مطالعات ماردیا و همکاران (۲۰۰۳) مدل دیگری به نام مدل کسینوسی با همان تعداد پارامتر در کارگاه آماري در دانشگاه لیدز پیشنهاد شد. سپس، ماردیا و همکاران (۲۰۰۷a) و و همکاران ماردیا (۲۰۰۷b) خواص متنوعی از این مدل را مطالعه کردند. در ادامه این فعالیت‌ها، کنت و همکاران (۲۰۰۸) به معرفي مدل دورگه[۲] از روي مدل‌هاي سينوسي و کسينوسي پرداختند. در همان زمان، ماردیا و همکاران (۲۰۰۸) تعمیم چند متغیره مدل سینوسی را ارائه دادند. اخیراً، بررسي تعمیم چندمتغیره این توزیع توسط ماردیا و واس (۲۰۱۴) صورت گرفت. در کنار مسائل متنوع نظري احتمالاتي مرتبط با چنبره، مطالعه مدل های آماري و نیز مورد توجه محققین قرار گرفته است. به عنوان مثال، اخیرا کنت و ماردیا (۲۰۱۵) درباره تعداد پیچش منحني‌ها روي چنبره مطالعه کردند تا بتوانند از آن براي معرفي مدل آماري رگرسیوني مناسب استفاده کنند. با این حال، ويژگيهاي ديگري از این مدلها شایسته تحقیق و توسعه هستند که در پژوهش پیشنهادي حاضر به یکي از آنها پرداخته خواهد شد.

مطالعه زوایای دوسطحی[۳] با رویکرد مدل‌بندی در پیشگویی ساختار پروتئین‌ها از اهمیت زیادی برخوردار است. از سالیان قبل، دانشمندان علوم زیستی پراکندگی نقاط مشاهده شده بر روی چنبره را در یک صفحه رسم می‌کردند که به افتخار پیشنهاد آن توسط راماچاندران و همکاران (۱۹۶۳) به آن، نمودار راماچاندران[۴] گفته می‌شود. در واقع نمودار راماچاندران وضعیت مجاز هر زاویه را برای ساختارهای پروتئین نشان می‌دهد. اگرچه نمایش تغییرات زوایا توسط نمودار راماچاندان از نظر شهودی آسان است، محققین را در مدلبندي احتمالاتی آن دچار اشتباه می کند. به عبارتی دیگر، توجه به نمایش واقعی نقاط روی چنبره می تواند انگیزه ای برای تفکر بیشتر در پایه ریزی مفاهیم متنوع آماری مانند رگرسیون بر روی این شکل هندسي شود. لذا، شکل ۱ (الف) نمودار راماچاندران پروتئینی خاص را نشان می‌دهد.

مقاله حاضر به اینصورت تدوین شده است که در بخش بعد جزئیاتی از توزیع فون میزس دومتغیره تشریح می شود. مدل بندي آمیختگی از توزیع فون میزس دومتغیره برای نیل به خوشه بندی در بخش ۳ می آید. تحلیل یک مثال واقعی مرتبط با مطالب مقاله در بخش ۴ خواهد آمد.

۲ توزیع فون‌میزس دومتغیره

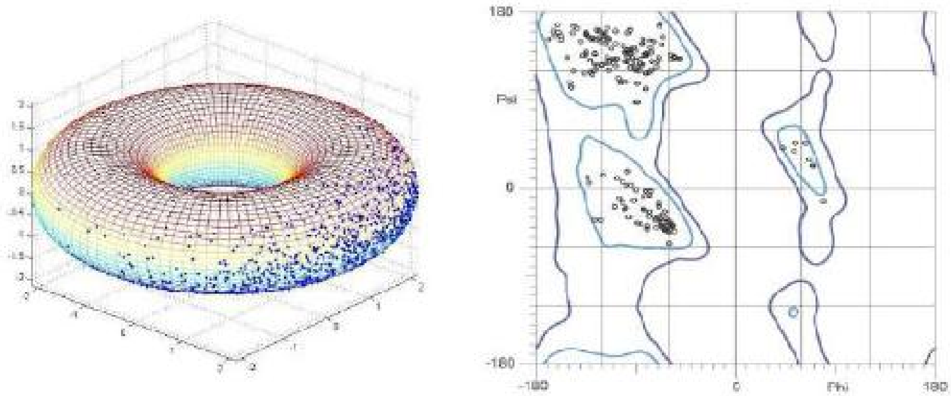
توزیع فون‌میزس یکي از توزیع‌هاي مفید و کارا براي مدل‌بندي داده‌هاي زاویه‌اي است. تحقیقات وسیعي راجع به حالت یك متغیره این توزیع صورت گرفته است که ماردیا و جاپ (۲۰۰۰) منبع مناسبی برای این منظور است. حالت دومتغیره آن اخیرا مورد توجه محققین آمار کاربردي بویژه آماردانان فعال در حوزه بیوانفورماتیك قرار گرفته است. ویژگی‌های مدل خاصی از حالت دومتغیره این توزیع، نحوه برآورد پارامترها و الگوی تولید نمونه از آن توسط نوری و گل علی زاده (۱۳۹۴) صورت گرفته است.

[۱] Bivariate Von mises

[۲] Hybrid Model

[۳] Dihedral angles

[۴] Ramachandran



شکل ۱: (الف): نمودار راماچاندران زوایای دوسطحی (ب):نمایش زوایای دو سطحی بر روی چنبره

دو زاویه θ_1 و θ_2 را طوریکه $-\pi \leq \theta_1, \theta_2 \leq \pi$ به‌عنوان دو متغیر زاویه‌ای درنظر بگیرید. حالت کلی تابع چگالی فون‌میزس دومتغیره که توسط ماردیا (۱۹۷۵) معرفی شده به‌صورت

$$\begin{aligned} f(\theta_1,\theta_2) &= (c(k_1,k_2,\mathbf{A}))^{-1}\exp\Big\{\kappa_1\cos(\theta_1-\mu_1)+\kappa_2\cos(\theta_2-\mu_2) \\ &+ [\cos(\theta_1-\mu_1),\sin(\theta_1-\mu_1)]\mathbf{A}[\cos(\theta_2-\mu_2),\sin(\theta_2-\mu_2)]^{\mathrm{T}}\Big\} \end{aligned}$$

(۱.۲)

است. به‌طوری که $\kappa_1, \kappa_2 > 0$ پارامترهای تمرکز و $-\pi \leq \mu_1, \mu_2 \leq \pi$ میانگین زاویه‌ای برای متغیرهای متناظر مربوطه هستند و $c(k_1,k_2,\mathbf{A})$ ثابت نرمال ساز است. در معادله (۱.۲) ماتریس $\mathbf{A}=[a_{ij}]$ ماتریسی با بعد 2×2 است. اگر در ماتریس $\mathbf{A}$ قرار دهیم $a_{12}=a_{21}=0$ و $a_{22}=\beta$ و $a_{11}=\alpha$ آنگاه داریم:

$$\begin{aligned} f(\theta_1,\theta_2) &= (c(k_1,k_2,\mathbf{A}))^{-1}\exp\Big\{\kappa_1\cos(\theta_1-\mu_1)+\kappa_2\cos(\theta_2-\mu_2) \\ &+ \alpha\cos(\theta_1-\mu_1)\cos(\theta_2-\mu_2) \\ &+ \beta\sin(\theta_1-\mu_1)\sin(\theta_2-\mu_2)\Big\}. \end{aligned}$$

حال با درنظر گرفتن $\beta=\lambda$ و $\alpha=0$ تابع چگالی توزیع فون‌میزس دومتغیره به‌صورت

$$\begin{aligned} f_{sin}(\theta_1,\theta_2) &= (c(k_1,k_2,\lambda))^{-1}\exp\Big\{k_1\cos(\theta_1-\mu_1)+k_2\cos(\theta_2-\mu_2) \\ &+ \lambda\sin(\theta_1-\mu_1)\sin(\theta_2-\mu_2)\Big\} \end{aligned} \qquad (۲.۲)$$

خواهد شد که سینگ و همکاران (۲۰۰۲) آن را مدل سینوسی نامیدند و ثابت نرمال‌ساز را به‌صورت زیر بدست آوردند:

$$C_1 = c(k_1,k_2,\lambda) = 4\pi^2\sum_{m=0}^{\infty}\binom{2m}{m}\left(\frac{\lambda^2}{4k_1k_2}\right)^m I_m(k_1)I_m(k_2).$$

واضح است که بنا به (۲.۲) اگر $\lambda = 0$ آنگاه θ_1 و θ_2 از هم مستقل می‌شوند و علاوه بر آن هرکدام از متغیرهای زاویه‌ای θ_1 و θ_2 بصورت جداگانه دارای توزیع فون‌میزس تک‌متغیره خواهند بود.

اگر $\alpha = \beta = -\kappa_3$ آنگاه تابع چگالی مدل کسینوسی توزیع فون‌میزس دومتغیره بصورت

$$\begin{aligned} f_{cosin}(\theta_1, \theta_2) &= (c(k_1, k_2, k_3))^{-1} \exp\Big\{k_1 \cos(\theta_1 - \mu_1) + k_2 \cos(\theta_2 - \mu_2) \\ &- k_3 \cos(\theta_1 - \mu_1 - \theta_2 + \mu_2)\Big\} \end{aligned} \qquad (۳.۲)$$

بدست می‌آید که در آن

$$\begin{aligned} C_2 &= c(k_1, k_2, k_3) \\ &= (2\pi)^2 \left[I_0(k_1) I_0(k_2) I_0(k_3) + 2\sum_{p=1}^{\infty} I_p(k_1) I_p(k_2) I_p(k_3)\right] \end{aligned}$$

ثابت نرمال ساز است و $k_1 \geq k_3$ و $k_2 \geq k_3$ و κ_3 پارامتر همبستگی است. واضح است که اگر در این مدل $\kappa_3 = 0$ آنگاه θ_1 و θ_2 از هم مستقل خواهند بود. علاوه بر این، اگر $\kappa_1 = \kappa_2 = \kappa_3 = 0$ آنگاه توزیع هر کدام از زوایا به توزیع یکنواخت روی دایره تبدیل می‌شود.

۳ مدل آمیخته توزیع فون‌میزس دومتغیره برای خوشه‌بندی زوایا

قابل ذکر است که در مدل های سینوسی و کسینوسی معرفی شده احتمال انتخاب نقاط بر روی چنبره برای تمامی نقاط یکسان در نظر گرفته شد. اما معرفی یک مدل آمیخته از این توزیع می تواند به احتمال انتخاب زوایای موجود در چنبره وزن‌های متناسب با تراکم حضورشان دهد که خود مبنای مناسبی برای خوشه بندی مجموعه ای از زوایا خواهد بود. به زبانی ساده‌تر، ممکن است نمودار راماچاندران برای زوج زوایای پیوندی پروتئین در تمام صفحه پراکنده نباشد و انباشتگی در بعضی از نقاط صفحه بیشتر از جاهای دیگر باشد. برای حل این مشکل ماردیا و همکاران (۲۰۰۷a) مدل آمیخته از توزیع فون میزس دومتغیره بصورت

$$f_M(\phi, \psi, \Theta) = \sum_{j=1}^{K} \pi_j f_j(\phi, \psi, \theta_j) \qquad (۱.۳)$$

پیشنهاد دادند. که در آن f_j تابع چگالی احتمال توزیع فون میزس دومتغیره برای پارامترهای $\theta_j = (k_{jk}, k = 1, 2, 3, \mu_{1j}, \mu_{2j})$ و $j = 1, 2, \ldots, K$ برای مدل کسینوسی و $\theta_j = (k_{jk}, k = 1, 2, \lambda_j, \mu_{1j}, \mu_{2j})$ برای مدل سینوسی هستند. همچنین $\pi_1, \pi_2, \ldots, \pi_K$ وزن هر تابع چگالی است ($\sum \pi_i = 1$). واضح است که برآورد پارامترهای این توزیع آمیخته از طریق روش مستقیم ممکن نیست. مرسوم ترین روش برای این حال، بکارگیری الگوریتم EM است. آنچه که معمول است تشریح گام های این الگوریتم است.

گام های الگوریتم تا رسیدن به همگرایی در زیر آمده است.

گام اول : برآورد اعضای احتمال با استفاده از

الف)

$$p_{ij} = \pi_j f_j(\phi_i, \psi_i, \theta_j)$$

برای هر $i = 1, 2, \ldots, n$ و $j = 1, 2, \ldots, K$

ب) نرمال کردن $p_{ij} = p_{ij} / \sum_j p_{ij}$ برای $i = 1, 2, \ldots, n$

گام دوم: برآورد ماکزیمم درستنمایی $\theta_j = (k_{jk}, k = 1, 2, \lambda_j, \mu_{j1}, \mu_{j2})$ که با ماکزیمم کردن تابع درستنمایی وزنی

$$\prod_{i=1}^{n} p_{ij} f_j(\phi_i, \psi_i, \theta_j)$$

برای $j = 1, 2, \ldots, K$ بدست می‌آید.

گام سوم: بدست آوردن نسبت آمیخته شدن $\pi_i = \sum_j p_{ij}$.

بنابر ماردیا و همکاران (۲۰۰۷a) انتقال داده ها $(\phi, \psi) \longrightarrow (\psi, -\phi)$ در مدل سینوسی توزیع فون‌میزس دومتغیره سبب همسانگردی با تغییر علامت پارامتر λ همراه است در حالی که در مدل کسینوسی باعث چرخش در نمودار کانتور می‌شود. این انتقال برازش بهتری برای این زوایای تصادفی فراهم می‌کند و با استفاده از این انتقال در گام دوم برآورد درستنمایی برای هر j بدست می آید. سپس، تعداد خوشه با استفاده از معیار اطلاعاتی آکائیکAIC بدست می‌آید.

۴ مطالعه یک مثال واقعی

برای تهیه مجموعه داده‌ها از پروتئین‌هایی که توسط NMR تعیین ساختار شده‌اند و تعداد متغیرهای آن برابر با چهل متغیر بود استفاده شد. برای بدست آوردن اطلاعات ساختار دوم از برنامه DSSP استفاده و بعد در محیط مطلب اسیدهای آمینه موجود در ساختارهای پروتئینی بر اساس اطلاعات ساختار دوم در سه دسته قرار گرفتند. این پروتئین متشکل از تعدادی اسید آمینه و هر اسید آمینه با زوایاي دوسطحي ϕ و ψ متصل می شوند. در پروتئین، ترتیب باقیمانده‌های اسید آمینه‌ها در زنجیره اصلی، تعیین‌کننده‌ی ساختار اول و تاخوردگی محلی [۵] زنجیره‌ای از اسیدآمینه‌ها بیانگر ساختار دوم آن است. آلفا هلیکس[۶] و صفحات بتا بخش عمده‌ای از ساختارهای دوم را در پروتئین تشکیل می‌دهند. اما این ساختارهای منظم از طریق نواحی که ساختار منظمی ندارند به هم متصل می‌شوند که آنها را لوپ[۷] یا کویل[۸] می نامند. کویل‌ها نواحی هستند که معمولاً در سطح پروتئین قرار می‌گیرند. این نواحی معمولا به صورت متصل‌کننده‌هایی بین دو ساختار منظم به حساب می‌آیند، اما می‌توانند از اهمیت ساختاری خاصی برخوردار باشند و حتی جایگاه فعال یا بخشی از عملکرد پروتئین را تشکیل دهند. در ساختار پروتئین مورد بررسی صفحات بتا و پیچهاي آلفا وجود دارند که توسط کویل به هم متصل شده‌اند. خواننده علاقه مند برای اطلاع بیشتر می تواند به راماچاندران و همکاران (۱۹۶۳) مراجعه کند.

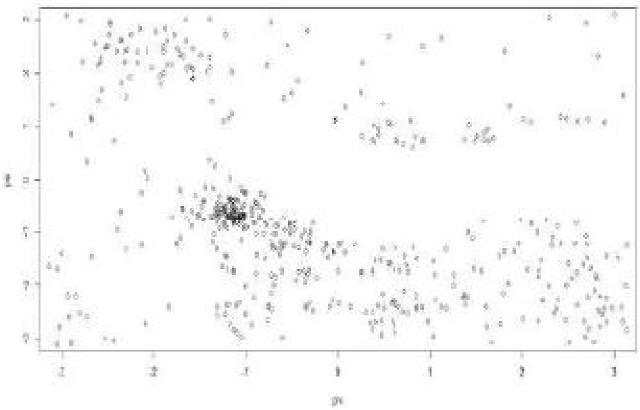
برای شروع تحلیل آزمون همبستگي براي داده‌هاي بتا و مارپیچ آلفا و کویل بکار برده شد و مقدار آماري آزمون و نتایج حاصل از آن در جدول ۱ آمده است. لازم به ذکر است که فرضیه صفر در این آزمون عدم وجود همبستگی بین متغیرهای مورد بررسی است. با نگاهی به $p-$ مقدار در جدول (۱) که برابر با ۱/۰۵۶ و ۰/۲۰۸ بدست آمده است می‌توان دریافت که فرضیه صفر در سطح ۰/۰۵ رد نمی‌شود و لذا مجموعه زوایاي دوسطحي همبستگي معني‌داري ندارند و در این صورت می‌توان نمونه‌ها را مستقل از هم در نظر گرفت. از نقطه نظر شهودی، به دلیل اینکه زوایای مورد مطالعه بر روی چنبره قرار دارد، از بین توزیع های موجود، توزیع فون میزس دومتغیره

[۵] local fold

[۶] Helix

[۷] Loop

[۸] Coil



شکل ۲: نمودار راماچاندران داده‌ها پروتئین

جدول ۱: نتایج آزمون همبستگی

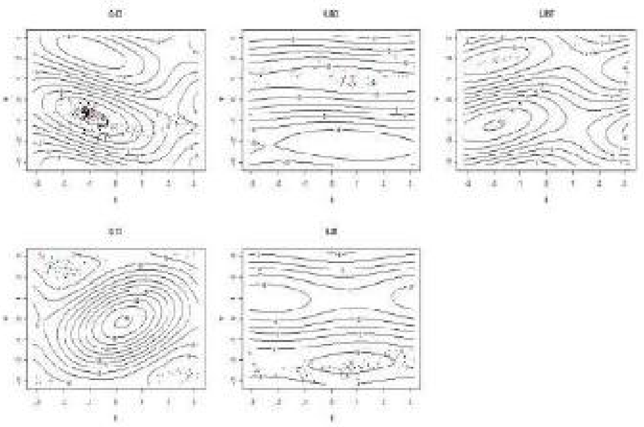
زاویه	$p-$ مقدار	مقدار همبستگی	مقدار آماره آزمون
ϕ	۰/۳۹۰	۰/۰۵۹	۱/۰۵۶
ψ	۰/۸۳۳	۰/۰۱۲	۰/۲۰۸

برای برازش مناسب‌تر بنظر می‌رسد. بنا به مطالب این مقاله و براساس نمایش راماچاندران مجموعه داده‌ها در شکل (۲)، مدل آمیخته‌ای از توزیع فون‌میزس دومتغیره برای تحلیل بیشتر داده‌ها مدنظر قرار گرفت. با برازش مدل ۳-۱ توسط الگوریتم EM پارامترهای توزیع برآورد شدند که و نتایج عددی حاصل در جدول (۲) آمده است. برای نمایش هندسی مولفه ها در توزیع آمیخته، شکل (۳) رسم شد. تعداد خوشه های اولیه برای این الگوریتم ۵ خوشه برای داده‌های هلیکس، بتا و کویل درنظر گرفته شد که مقدار احتمال ۳ خوشه بیشتر از همه بوده و بنابراین تعداد خوشه ها همانگونه که از ابتدا انتظار می‌رفت ۳ خوشه هلیکس، بتا و کویل نشان داده می شود.

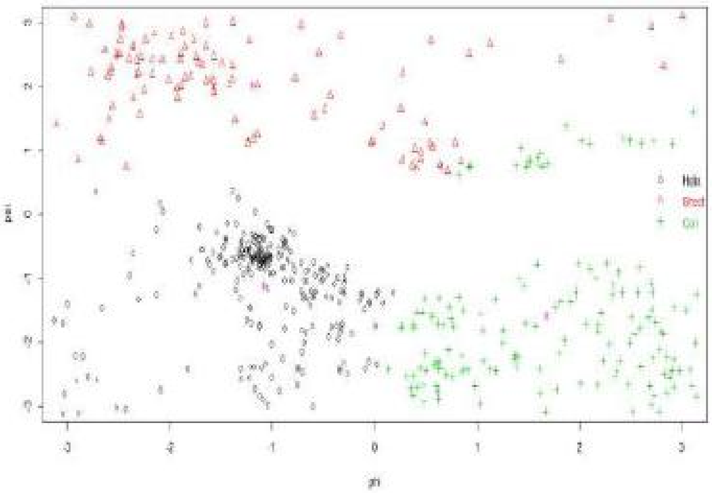
جدول ۲: برآورد EM پارامترهای مدل آمیخته

خوشه	κ_1	κ_2	κ_3	μ_1	μ_1	π
۱	۶/۲۳	۰/۳۴	−۳/۹	−۰/۸	۰/۹۳	۰/۴۳
۲	۱۱/۵۳	۱۵/۰۳	۸/۹۹	−۲/۸۵	۲/۹۹	۰/۱۵
۳	۶/۰۳	۱/۰۶	۰/۳۹	−۲/۱۳	−۰/۷۴	۰/۲۸

با فرض اینکه خوشه‌بندی مناسب این داده ها بوده و تعداد خوشه موجود ۳ هست، ترجیح داده شد از روش های موجود در خوشه بندی استفاده شود. روش $k-means$ برای خوشه‌بندی این داده ها استفاده شد که درصد دقت روش خوشه‌بندی برابر با ۶۷/۸ بدست



شکل ۳: نمودار کانتور برای هر مولفه در مدل آمیخته داده‌های پروتئین



شکل ۴: نمودار داده‌ها پس از خوشه‌بندی

آمد. به دلیل اینکه مطالعه پروتئین ها براساس مدل های مانند آنچه که در این مقاله از رویکردهای نوین تحقیق آماری است، به نظر می رسد درصد دقت بدست آمده تاحدی قابل قبول باشد. با این حال تحقیق بیشتر در این زمینه ضروری است. در شکل (۴) نحوه خوشه بندی داده ها نیز رسم شده است. به عنوان مثال، می توان از معیارهای فاصله دیگری بر روی خمینه‌ها مانند فاصله ژئودزیک برای بهبود این الگوریتم نیز بهره گرفت که این موضوع به عنوان هدف آتی در تحقیق مدنظر نویسندگان است.

مراجع

نوری جویباری، س. و گل‌علی‌زاده، م. (۱۳۹۴)، برآورد شبه‌درستنمایی توزیع فون‌میزس دومتغیره، مجله مدل سازی پیشرفته ریاضی، دانشگاه شهید چمران اهواز، ۴، ۸۶-۷۱.

Kent, J. T., Mardia, K. V., and Taylor, C. C. (2008), Modelling strategies for bivariate circular data. In

Proceedings of the Leeds Annual Statistical Research Conference, The Art and Science of Statistical Bioinformatics, Leeds University Press, Leeds , pp. 70-73.

Kent, J. T. and Mardia, K. V. (2015), The winding Number for Circular Data, In *Geometry-Driven Statistics and its Cutting Edge Applications: Celebrating Four Decades of Leeds Statistics Workshop.*, Eds. K.V. Mardia, A. Gusnanto, C. Nooney and J. Voss. Leeds University Press, 47-50.

Krishnan, T. H. R. I. Y. A. M. B. A. K. A. M., and McLachlan, G. J. (1997), *The EM algorithm and extensions*, John Wiley, New Jersey.

Mardia, K. V. (1975). Statistics of Directional Data, *Journal of the Royal Statistical Society, Series B*, **37**, 349-393.

Mardia, K. V. and Jupp, P. (2000). *Directional Statistics*, John Wiley, New York.

Mardia, K. V., Taylor, C. C. and Subramanian, G. K. (2003), Applications of Circular Distribution to Conformational Angles in Proteins, In *Proceedings of the 22nd LASR Workshop*, edited by KV Mardia, RG Aykroyd and MJ Langdon, 149-152. Leeds University Press.

Mardia, K. V., Taylor, C. C. and Subramanian, G. K. (2007a), Protein Bioinformatics and Mixtures of Bivariate Von Mises Distribution for Angular Data, *Biometrics*, **63** , 502-512.

Mardia, K. V., Taylor, C. C. and Subramaniam, G. K. (2007b), Bivariate Von Mises Densities for Angular Data with Application to Protein Bioinformatics, *Annals of Statistics*, **35**, 166-180.

Mardia, K.V., Hughes, G., Taylor, C.C. and Singh, H. (2008), A Multivariate Von Mises Distribution with Applications to Bioinformatics, *Canadian Journal of Statistics*, **36**, 99-109.

Mardia, K. V. and Voss, J. (2014), Some Fundamental Properties of a Multivariate Von Mises Distribution, *Communications in Statistics-Theory and Methods*, **43**, 1132-1144.

Ramachandran, G. N., Ramakrishnan, C., Sasisekharan, V. (1963), Stereochemistry of Polypeptide Chain Configurations, *Journal of Molecular Biology*, **7**, 95-99.

Singh, H., Hnizdo, V. and Demchuk, E. (2002). Probabilistic Model for Two Dependent Circular Variable, *Bioinformatics*, **89**, 719-723.