فهم ساختار موضوعی قرآن کریم: روی‌کرد چند متغیّرهٔ اکتشافی ـ
نجلا ثابت، سیّد مصطفی احمدزاده، مهدی جباری نوقانی
فصلنامه تخصّصی مطالعات قرآن و حدیث **سفینه**
سال پانزدهم، شماره ۶۰ «ویژه قرآن‌بسندگی»، پاییز ۱۳۹۷، ص ۱۱۱-۱۳۰

فهم ساختار موضوعی قرآن کریم: روی‌کرد چند متغیّرهٔ اکتشافی

نجلا ثابت*
سیّد مصطفی احمدزاده**
مهدی جباری نوقابی***

چکیده: در این مقاله، روشی‌شناسیِ کشف ساختار موضوعی قرآن کریم بر اساس یک ایدهٔ بنیادی در داده‌کاوی و شیوه‌های مرتبط با آن ارائه شده و آن این که در برخی از متون، فراوانی گونه‌های واژه‌ها، شاخص[1] معتبری برای محتوای مفهومی آن‌ها است و این ویژگی، معیار معتبری برای طبقه‌بندی آن واژه‌ها نسبت به یکدیگر فراهم می‌سازد. از این روش، برای کشف ارتباطات متقابل موضوعی میان سوره‌های قرآن کریم از طریق لحاظ نمودن فراوانی داده‌های واژه‌ایِ آن سوره‌ها و سپس به کاربردن تحلیل خوشه‌ای به شیوه سلسله مراتبی، استفاده می‌شود. نتایجی که در اینجا گزارش شده، نشان می‌دهد که روش‌شناسی حاضر نتایج کاربردی[2] در فهم قرآن کریم بر اساس

*. دانشکده زبان، ادبیات و زبانشناسی انگلیسی، دانشگاه نیوکاسل n.a.Thabet@ncl.ac.uk
**. استادیار پژوهشگاه علوم و فرهنگ اسلامی M.ahmadzadeh@isca.ac.ir
***. استادیار گروه آمار، دانشگاه فردوسی مشهد، مشهد، ایران jabbarinm@um.ac.ir

۱. برای دستیابی به محتوای مفهومی واژه‌ها، شاخص‌های گوناگونی باید مد نظر قرار گیرد. در این میان، نخستین شاخص و یکی از شاخص‌های کمّی در این ساحت که به زیرساختی برای شاخص‌های کیفی به شمار می‌آید، بررسی فراوانی گونه‌های واژه‌هاست. (مترجم)

۲. هم‌چنان که مؤلف تأکید کرده، این روش، یک روش کاربردی است و نتایج آن نیز سمت و سوی کاربردی دارد. لذا نباید از این روش انتظار داشت که به دایره‌ی روش‌های بنیادی وارد شود. به دیگر سخن، این روش، زمینه را برای روش‌های بنیادی فراهم می‌کند و مقدمه‌ای برای ورود به روش‌های بنیادی به شمار می‌آید. (م.)

معنی‌شناسی[1] واژه‌های قرآن کریم دارد.

کلیدواژه‌ها: تفسیر موضوعی؛ روش‌شناسی ترکیبی؛ ساختارشناسی سوره‌ها؛ داده‌کاوی قرآنی.

۱. مقدمه

قرآن کریم یکی از کتاب های بزرگ مذهبی دنیا است و در قلب فرهنگ اسلامی جای دارد. تفسیر جامع و دقیق قرآن کریم، هم برای اعتقاد میلیون‌ها مسلمان در سراسر جهان و هم برای فهم دنیای غیر اسلام از دین اسلام، مهم است. سنّت قدیمی تفسیر عالمانهٔ قرآن کریم، ضرورتاً مبتنی بر روش‌های تاریخی- ادبیِ سنّتیِ تفسیرِ متن است[2] که به صورت دستی (و نه ماشینی) صورت می‌گیرد. امّا، در حال حاضر، پیشرفت‌ها در ارائه و تحلیل متن به صورت الکترونیکی، امکان به کارگیری فناوری‌های موجود در حوزه‌های پژوهشی نوپدید مانند داده‌کاوی را در مورد تفسیر قرآن کریم فراهم می‌سازد. تقریباً در حوزهٔ مطالعات تحلیل آماری قرآن کریم، کاری انجام نشده است.[3] نگارش مقالات در این رشته، موجب بسط و توسعهٔ

۱. در معنی‌شناسی زبانی به آن گونه از معنی توجه می‌شود که درون زبانی است و واحدهای مطالعه‌ی این نوع معنی نیز واژه و جمله به حساب می‌آیند (صفوی، ۱۳۷۹، ٤٦). (م.)

۲. هر چند این نکته در جای خود درست است، اما در میراث ماندگار دانش تفسیر، قراینی وجود دارد که نشان می‌دهد دانشمندان اسلامی از توجه به بحث‌های آماری در قرآن کریم غفلت نورزیده‌اند. دانش «وجوه و نظائر»، روش «تفسیر قرآن به قرآن»، «فرهنگ‌های تخصصی قرآن کریم»، بررسی و نقد آراء مذاهب گوناگون درباره‌ی برداشت‌های مختلف از آیات قرآن کریم و ذکر دقایق و ظرایف عرفا در تفسیر عرفانی قرآن کریم، همگی شاهدی بر این مدعا است. قرآن‌پژوهان برای ورود به این بحث‌ها از آمارهای قرآنی به عنوان زمینه و مقدمه‌ی اثبات دیدگاه خویش بهره می‌برده‌اند؛ هر چند این آمارها به دقت و جامعیّت امروزی نبوده و نیز تنها با استفاده از روش های آمار توصیفی انجام گرفته است. در این مقاله نویسنده محترم با استفاده از روش‌های آمار استنباطی توام با آمار توصیفی نتایج ملموس و جالبی بدست آورده است. (م.)

۳. اگر نویسنده‌ی محترم به جای جمله «تقریباً ...» از عبارت «در حوزه‌ی مطالعات تحلیل آماری قرآن کریم، کارهای اندکی انجام شده است» استفاده می‌کرد، به صواب نزدیک‌تر بود. زیرا در ساحت تحلیل آماری قرآن کریم، پژوهش‌های قابل توجهی در دهه‌های اخیر توسط دانشمندان مسلمان انجام گرفته است که برخی از آن‌ها عبارت است از: روحانی (۱۳۶۸)؛ بازرگان (۱۳۵۵)؛ Khalifa(1982)؛ رفاعی (۱٤۲۱)؛ جرار (۱٤۱٤)؛ دیدات (۱۳۸۱) و نوفل (۱۳۸۳). برای نقدهای اعجاز عددی و آثار اخیر ر. ک : یزدانی، ۱۳۷۵، ۸٤- ٦۲. (م.)

[DOR: 20.1001.1.17350689.1397.15.60.7.8]

نوعی تحلیل‌گر واژه‌شناسی برای قرآن کریم شده است.[1] (Dror et al., 2004)

قرآن کریم یکصد و چهارده سوره دارد که در پیوستاری از بلندترین سوره، بقره با دویست و هشتاد و شش آیه تا کوتاه‌ترین سوره، کوثر با چهار آیه مرتب شده است. دلیل روشنی در دست نیست که چرا سوره‌ها آن چنان که در قرآن کریم هستند، مرتب شده‌اند. این سوره‌ها ترتیب تاریخی ندارند و به نظر می‌رسد تقریباً بر اساس طول از بزرگ‌ترین سوره در ابتدای قرآن کریم تا کوچک‌ترین سوره در انتهای قرآن کریم مرتب شده‌اند. با این فرض که مرتب‌سازیِ این سوره‌ها ظاهراً اجتهادی بوده، یکی از گام‌های نخستین در تفسیر قرآن کریم به عنوان یک کلّ، کشف روابط موضوعی میان این سوره‌ها است.[2]

این مقاله، روش‌شناسی‌ای برای انجام این‌گونه مطالعات، از طریق تحلیل چند متغیرهٔ اکتشافی ارائه می‌دهد.

۱. منظور نویسنده این است که در تحلیل‌های آماری واژه‌های قرآن کریم تاکنون تنها روش‌های آمار توصیفی همانند جداول فراوانی، شاخص‌های آماری تمایل مرکزی و پراکندگی (همانند میانگین و... و واریانس و ...) و نمودارهای آماری بکار گرفته شده است و به کاربرد بردن روش‌های آمار استنباطی همانند مقایسه‌های شاخص‌ها و بخصوص روش‌های مدل‌سازی و خوشه-بندی و ... یا اصلا استفاده نشده یا کمتر مورد اقبال بوده است.

۲. با توجه به رشته‌ی تخصصی نویسنده‌ی محترم که زبانشناسی است، طبیعی است که اطلاعات او درباره‌ی مباحث قرآنی کم باشد. اما هر پژوهشگری باید اطلاعات خود را نه تنها در ساحتی که پژوهش می‌کند؛ بلکه در حوزه‌های میان‌رشته‌ای مربوط به پژوهش خود نیز تقویت کند تا بر استحکام، قوت و ارزش علمی کار پژوهشی خویش بیفزاید. دانشمندان اسلامی درباره‌ی ترتیب سوره‌ها در قرآن کریم در طول هزار و چهار صد سال گذشته بسیار بحث کرده‌اند و نظریات گوناگونی نیز در این زمینه ارائه داده‌اند. این نظریات در سه دسته‌ی کلی قابل تقسیم است: توقیفی بودن ترتیب سوره‌ها، اجتهادی بودن ترتیب سوره‌ها، قول به تفصیل که چینش عمده‌ی سور را توقیفی می‌داند و اجتهاد و نظر صحابه را تنها در تنظیم چند سوره، دخیل معرفی می‌کند. در میان این دیدگاه‌ها، دیدگاه تفصیلی از اعتبار بیشتری برخوردار است و مستندات آن نسبت به سایر دیدگاه‌ها از قوت بیشتری برخوردار است. (جلالی نائینی ۱۳۸٤، ۱۵٦-۱۵۳) بر اساس پذیرش دیدگاه نخست و سوم، بحث تناسب میان سور قرآن کریم برجسته می‌شود. قرآن‌پژوهان و مفسّران در آثار و تفاسیر خود، در انتها و ابتدایِ هر سوره، بیشتر به ارتباط میان دو سوره‌ای که در کنار یکدیگر قرار گرفته‌اند، پرداخته‌اند و کمتر به ارتباط میان سوره‌های قرآن کریم به عنوان یک مجموعه‌ی منسجم پرداخته‌اند. هر چند در این مورد نیز آثاری پدید آمده است که عبارتند از: سیوطی (۲۰۰۰) و الهی‌زاده (۱۳۹۱). (م)

این مقاله چهار[1] بخش دارد: بخش نخست، درآمد است. بخش دوم، متن قرآنی و آمایش داده‌ها پیش از تحلیل را ارائه می‌دهد. بخش سوم، به کاربرد فنون تحلیل خوشه‌ای در مورد قرآن کریم و تفسیر نتایج آن می‌پردازد. بخش چهارم، نتیجه‌گیری و موضوعاتی را برای پژوهش‌های بعدی پیشنهاد می‌کند.

۲. داده‌ها

داده‌های این تحقیق بر اساس نسخه‌ی الکترونیکی قرآن کریم محصول شرکت «مسلم‌نت» است. این نسخه، حرف‌نگاری الفبای غربی، از املاء عربی است. این داده‌ها بر اساس علائم[2] ASCII، غالباً با معادل‌های تک - نماد از آواهای عربی و با جایگزینی اِعراب و نشانه‌هایی که واکه‌های کوتاه را در املاء عربی با حروف لاتینی متناسب با آن‌ها نشان می‌دهند، لاتین نگاری شده است.

در ماتریس فراوانی F، سوره‌ها در ردیف‌ها و واژه‌ها در ستون‌ها ارائه شده‌اند و در هر سلول F_{ij}، یک عدد کامل قرار دارد که تعداد دفعات واژهٔ j را که در هر سورهٔ i واقع شده، نشان می‌دهد. اصولاً طراحی این‌گونه ماتریس‌ها آسان است، اما در عمل برخی مسائل شناخته شده بروز می‌کند.

۲-۱. رمزگشایی

اگر شخصی بخواهد واژه‌ها را بشمارد، نخستین سؤالی که برای او پیش می‌آید این است که واژه چیست؟ پاسخ به این سؤال به صورت شگفت‌آوری مشکل است و این مشکل، هم در زبان شناسی و هم در پردازش زبان طبیعی وجود دارد. (مانینگ و شوتز، ۱۹۹۹)[3] حتی اگر مانند این مقاله، محدود به زبان نگارش گردد، دو

۱. در اصل مقاله، پنج آمده است که ظاهراً در ویرایش نهایی نویسنده‌ی محترم تصحیح نشده است. (م.)

۲. مهم‌ترین استاندارد برای نمایش حروف، اعداد و علائم در نرم‌افزارهای رایانه‌ای است. (م.)

3. Manning & Shütze.

[DOR: 20.1001.1.17350689.1397.15.60.7.8]

اختلاف نظر رخ می‌دهد:

- تقطیع کلمات: در متون انگلیسی، این باور عمومی که واژه، رشته‌ای از حروف به هم چسبیده است که از طریق علائم نقطه‌گذاری و/ یا فضای خالی مشخص می‌شود، دیدگاه جا افتاده‌ای است؛ در حالی که در مورد زبان‌های دیگر این باور کم‌تر است.

- ستاک (بُن): در زبان‌هایی که بن‌های واژه‌های آن زبان‌ها دارای اشتقاقات صرفیِ گسترده است؛ آیا اشتقاقات گوناگون صرفیِ یک بُن خاص، به عنوان واژه‌های متفاوت به شمار می‌آیند؟

در قرآن کریم، تقطیع کلمه‌ها به راحتی قابل حل است، زیرا املاء قرآن کریم به گونه‌ای است که با استفاده از دو معیارِ علائم نقطه‌گذاری و/ یا فضای خالی، می‌توان واژه‌ها را به صورت قابل اعتمادی شناخت. بُن و اشتقاقات صرفی نیز به مانند گونه‌های یک واژه به شمار می‌آیند[1] و برای دستیابی به این موارد، متن الکترونیکی

۱. نویسنده‌ی محترم معتقد است که دو اختلاف بالا در مورد قرآن کریم، به دلیل ساختار زبان عربی و رسم الخط قرآن کریم وجود ندارد و یا دست کم قابل توجّه نیست؛ اما مشکلات دیگری درباره‌ی قرآن کریم بروز می‌کند که پرداختن به آن‌ها به عنوان مبانی نظری روش‌شناسیِ پیشنهاد شده ضروری است که متأسفانه در متن مقاله نیامده است. برخی از این نکات عبارتند از: ۱. بی‌توجهی به بافت‌های گوناگون زبانی واژه‌های قرآن کریم: روشن است که هر واژه‌ای در بافتی قرار دارد که بخش قابل توجهی از معنای خود را مرهون بافتی است که در آن قرار گرفته است. بدون توجه به بافت هر واژه، نمی‌توان معنای دقیقِ آن واژه را به دست آورد، زیرا یک واژه در دو بافت گوناگون، دو معنای متفاوت دارد. این در حالی است که در این روش‌شناسی، برای بررسی ارتباطات میان سوره‌های قرآن کریم، تنها به فراوانی کاربردهای واژه‌ها در سوره‌ها- بدون بررسی بافت هر واژه و در نتیجه معنای هر واژه- توجّه شده است. این گونه پژوهش‌ها به دلیل تداخل معانی واژه‌ها، از روایی (validity) لازم برخوردار نیستند. این نقیصه از جمله مواردی است که نویسنده‌ی این نوشتار باید پاسخ قانع‌کننده‌ای برای آن ارائه می‌کرد.

۲. بی‌توجهی به واژه‌های ترکیبی: در زبان عربی همانند هر زبانی، انبوهی از واژه‌ها به صورت ترکیبی به کار می‌روند. مثلاً واژه‌ی «ضرب» با ترکیب با واژه‌هایِ (الامثال، الارض، ...) معانی متفاوتی پیدا می‌کند (ر.ک: عمر، ۱۳۸۵، ۱۴۶)، که در صورت شمارش جداگانه‌ی واژه‌ها، از یک سو این معانی در داده‌های آماری راه پیدا نمی‌کنند و از سوی دیگر، آن چیزی مورد سنجش قرار می‌گیرد که واقعیّت ندارد. به دیگر سخن، پژوهش‌هایی که بر پایه‌یِ این روش‌شناسی انجام می‌شود، از روایی بالایی برخوردار نخواهند بود. این هم نکته‌ای است که نویسنده‌ی محترم مقاله پاسخی برای حل آن ارائه نکرده است. ۳. بی‌توجّهی به واژه‌های چند معنا: در زبان عربی و به ویژه در قرآن کریم، واژه‌هایی به کار رفته که چند معنا دارند. دانش «وجوه و نظائر»

[DOR: 20.1001.1.17350689.1397.15.60.7.8]

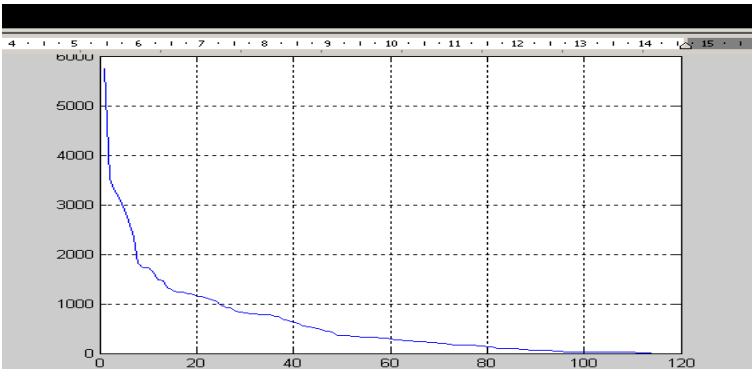
قرآن کریم با استفاده از بُن‌ساز محقق‌ساخته که مشخصات و عملکرد آن در مقالهٔ «ثابت» (۲۰۰۴) شرح شده، پردازش شد.

۲-۲. انتخاب واژه‌های کلیدی

واژه‌های نقش‌نما، مانند ضمایر اشاره و حروف اضافه از متن حذف شدند و تنها واژه‌های معنادار باقی ماندند. افزون بر این، واژه‌هایی که تنها یک بار به کار رفته بودند و در تعیین ارتباط میان سوره‌ها نقشی نداشتند، حذف شدند.

۲-۳. استانداردسازی برای طول متن

در مقدمه متذکر شدیم که طول سوره‌ها از حدود دوازده تا چند هزار کلمه در نوسان است. نمودار زیر، تعداد واژه‌های معنادار در هر سوره را -که براساس کاهش تعداد، مرتب شده- نشان می‌دهد.



شکل ۱: نمودار تعداد واژه‌های معنادار در هر سوره

← برای بحث درباره‌ی این واژه‌ها در قرآن کریم توسط لغت‌شناسان مسلمان بنیان نهاده شده است و آثار متعددی در این زمینه به رشته‌ی تحریر درآمده است. در این دانش از واژه‌هایی مانند «یوم» که در آیات گوناگون، معانی مختلفی دارد، بحث می‌شود. اگر تنها به فراوانی یک واژه- بدون توجّه به معانی متفاوت آن- اکتفا کنیم، دست‌کم نتایج این گونه تحلیل‌ها در مورد قرآن کریم صادق نخواهد بود. زیرا این گونه واژه‌ها-که تعداد آن‌ها در قرآن کریم نیز بسیار است- در سوره‌های گوناگون، معانی متفاوتی دارند، در صورتی که در این گونه تحلیل‌ها، این تفاوت‌ها مد نظر قرار نمی‌گیرد و بر پایه‌ی آن، نتایج به دست آمده با آن‌چه در قرآن کریم وجود دارد، فاصله خواهد داشت. از این رو، شایسته بود نویسنده‌ی محترم در خصوص رفع این مشکلات، راه کارهایی روش‌شناسانه ارائه دهد. نکات دیگری نیز در این‌باره قابل طرح است که به دلیل اختصار از ذکر آن‌ها خودداری می‌شود. (م.)

[DOR: 20.1001.1.17350689.1397.15.60.7.8]

روشن است چنان‌چه واژه‌ای با احتمال وقوع کم داشته باشیم، احتمال بیشتری دارد که این واژه در متنی طولانی بیاید تا در متنی کوتاه. بنابراین، به منظور مقایسهٔ معنادار سوره‌ها بر اساس نمودار واژه‌هایشان، می‌باید فراوانی‌های اولیه واژه‌ها را تعدیل نماییم تا بتوانیم نوسان طول سوره‌ها را جبران کنیم. این تعدیل، بر اساس فرمول زیر انجام شده است:

$$freq'(F_{ij}) = freq(F_{ij}) \times \left(\frac{\mu}{l}\right)$$

که در آن $freq'$ فراوانی تعدیل شده، F_{ij} مقدار مختصات داده‌های ماتریس F، $freq$ فراوانی اولیه، μ مقدار میانگین تعداد واژه‌ها در هر سوره در گسترهٔ همهٔ یکصد و چهارده سوره، و l تعداد واژه‌ها در سورهٔ $\mathbf{i}$ام است.

همچنین باید توجّه شود که به همان نسبتی که طول متن کوتاه می‌شود، احتمال وقوع یک واژهٔ مشخص در سوره، حتی برای یک بار، نیز کم می‌شود و بنابراین، بردار فراوانیِ آن نیز به صورت فزاینده‌ای تُنُک خواهد شد، که عمدتاً شامل صفر می‌باشد. زیرا صفر تعدیل‌ناپذیر است، توابعی که متغیر طول متن را خنثی می‌کنند به صورت فزاینده‌ای نتایج غیرقابل اعتماد تولید می‌کنند، در حالی که طول متن کاهش یافته است.[1] بنابراین، در مقالهٔ حاضر، تنها سوره‌های نسبتاً طولانی و به ویژه بیست و چهار سوره‌ای که هزار یا بیش از هزار واژهٔ محتوایی دارند، مورد ملاحظه قرار گرفته است.

۱. نویسنده‌ی محترم در این‌جا به برخی از کاستی‌های این روش- دست کم برای ارتباطات متقابل موضوعی میان سوره‌های بزرگ و کوچک - اذعان کرده است. (م.)

نام سوره	واژه‌ها	نام سوره	واژه‌ها	نام سوره	واژه‌ها	نام سوره	واژه‌ها
البقره	۵۷۳۹	الانفال	۱۱۵۶	الاسراء	۱۴۶۴	الشعراء	۱۲۰۸
آل عمران	۳۳۱۶	التوبه	۲۳۴۵	الکهف	۱۴۸۶	النمل	۱۰۶۹
النساء	۳۵۴۳	یونس	۱۷۳۲	طه	۱۲۶۵	القصص	۱۳۳۲
المائده	۲۶۸۱	هود	۱۸۰۹	الانبیاء	۱۰۷۷	الاحزاب	۱۲۳۹
الانعام	۲۸۹۵	یوسف	۱۶۶۵	الحج	۱۱۹۵	الزمر	۱۱۰۷
الاعراف	۳۱۲۷	النحل	۱۷۲۹	النور	۱۲۳۶	غافر	۱۱۵۶

جدول ۱: لیست سوره‌های با بیش از هزار واژه

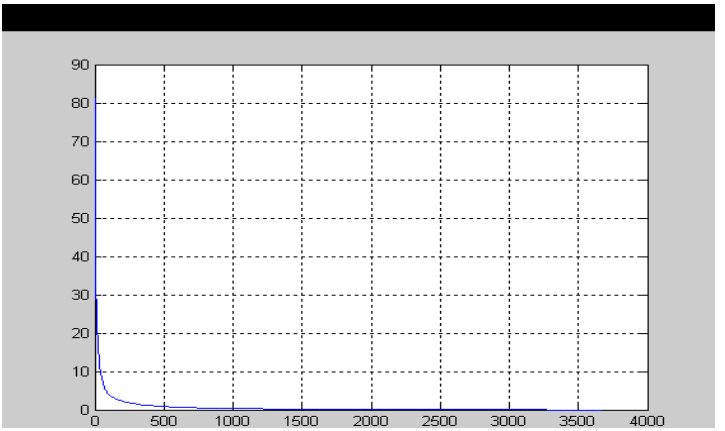
انتخاب هزار واژه به عنوان آستانهٔ طول، اختیاری است. این‌گونه انتخاب‌های اختیاری، به روش‌شناسی مقاله آسیبی نمی‌رساند. امّا روشن است که هرگونه تحلیل ضابطه‌مندی از قرآن که از این روش‌شناسی بهره می‌گیرد، می‌باید از پسِ این مشکل برآید که کدام سوره‌ها را به روش اصولی بر مبنای طولشان مستثنی نماید؛ البته اگر چنین مشکلی وجود داشته باشد.

۴-۲. کاهش بعد

پس از حذف واژه‌های نقش‌نما و واژه‌های با یک بار فراوانی و اشتقاقات صرفی، تعداد ۳۶۷۲ واژهٔ معنادار به لحاظ موضوع، باقی ماند که به یک ماتریس با ۳۶۷۲ ستون نیاز دارد. به فرض این که تنها بیست و چهار جزء داده داشته باشیم، این به داده‌هایی به شدت پراکنده منجر می‌گردد که ابعاد آن باید تا حد امکان کاهش یابد تا با نیاز به نمایش متناسب با حوزهٔ داده‌ها هماهنگ گردد. برای اطلاع بیشتر دربارهٔ روش‌های کاهش بعد داده‌ها، به ورلیسن[1] مراجعه کنید. (۲۰۰۳) بدین منظور، پراکندگی همهٔ ۳۶۷۲ ستونِ فراوانیِ ماتریس F محاسبه و به ترتیب درصد تغییراتی

1.Verleysen.

که از واریانس تبیین می‌شود به صورت نزولی مرتب و روی نمودار زیر نشان داده شده‌اند:



شکل۲: نمودار پراکنش برای ۳۶۷۲ ستون

این همان چیزی است که انتظار داریم توزیع فراوانی واژهٔ نمونه در متن زبان طبیعی از قانون زیپف پیروی کند. (مانینگ و شوتز، ۱۹۹۹: ۲۶-۲۰) تقریبا همهٔ پراکندگی داده‌ها توسط مجموعه کوچک‌تری همانند پانصد بُعد یا کمتر قابل تبیین می‌باشد. درصدی از پراکندگی که توسط باقی مانده بُعدها تبیین می‌شود، آن قدر کم است که نقش معناداری در تبیین واریانس نداشته و بنابراین می‌توان آن را نادیده گرفت. بنابراین این ماتریس به پانصد متغیر/ ستون محدود شده که به تحلیل خوشه‌ای ماتریس 500×24 منجر می‌شود.

۳. تجزیه و تحلیل

۳-۱. تحلیل خوشه‌ای سلسله مراتبی

هدف تحلیل خوشه‌ای آن است که غیر تصادفی بودن توزیع بُردارها را در یک فضایی از داده‌ها طوری مشخص نموده و به صورت نموداری نشان دهد که فاصلهٔ بین گروه‌ها ارتباط کمی با بُعد فضای داده‌ها داشته و فاصله درون گروهی همبستگی

بزرگ‌تری با بُعد فضای داده‌ها داشته باشد. اطلاعات مفصّل‌تری از آنالیز خوشه‌ای سلسله مراتبی در اوریت (۲۰۰۲)،[1] گوردن (۱۹۹۹، ۱۰۹-۶۹)[2] و گوری (۲۰۰۰)[3] آمده است. برای بحث‌های کوتاه‌تر به اوریت و دون (۲۰۰۱،۱۶۰- ۱۲۵)،[4] هایر و همکاران (۱۹۹۸،۵۱۸-۴۶۹)،[5] فلاین و دیگران (۱۹۹۹،۹-۲۷۵)،[6] کاچیگان (۱۹۹۱،۷۰-۲۶۱)[7] و اُاَکس (۱۹۹۸،۱۲۰-۱۱۰)[8] نگاه کنید. تحلیل خوشه‌ای، دو گونهٔ اصلی دارد: سلسله مراتبی و غیر سلسله مراتبی. هدف در اینجا تنها یافتن خوشه‌ها و نمایش آن‌ها به صورت نموداری نیست، بلکه همچنین ارتباطات اساسی میان اجزاء داده‌ها و خوشه‌های اجزاء داده‌ها را با رسم درختواره نگار[9] یا درخت‌ها باید نشان داد.

آنالیز خوشه‌ای سلسله مراتبی، تشابه میان بردارها را به عنوان پایه‌ای برای خوشه‌بندی به کار می‌برد، به طوری که می‌تواند نزدیکی را تحت عنوان تشابه یا تفاوت (دوری) اندازه‌گیری نماید. با درنظر گرفتن داده‌های این تحقیق، نزدیکی در واژه‌ها، یا واژه‌های مشابه یا واژه‌های غیرمشابه سنجیده می‌شود؛ البته اغلب تمایز مورد سنجش واقع می‌شود که در این جا نیز این حالت انتخاب شده است. ساخت درخت خوشهٔ مبتنی بر تمایز، برای ماتریس داده‌هایی که مقدارهای عددی دارند، مانند موردی که در بالا توصیف شده، دارای دو مرحله است. در گام نخست،

1. Everitt.
2. Gordon.
3. Gore.
4. Dunn & Everitt.
5. Hair.
6. Flynn.
7. Kachigan.
8. Oakes.
9. Dendrogram.

جدول تمایزهایِ میان اجزاء داده‌ها، یعنی میان بُردارهای سطری ماتریس داده‌ها، ساخته می‌شود. اندازه به کار گرفته شده اغلب فاصله اقلیدسی است؛ در این روش، فاصلهٔ میان بردارهای A و B بر اساس فرمول معروف زیر محاسبه شده است:

$$length(z) = \sqrt{\left(length(x)\right)^2 + \left(length(y)\right)^2}$$

اندازه‌های بسیار دیگری در Gorden (۱۹۹۹،۱۵-۳) و Flynn (۱۹۹۹،۲۷۴-۲۷۱) و دیگران، وجود دارد.

سپس در گام دوم، جدول تمایز، برای ساخت خوشه با الگوریتم عمومی زیر، مورد استفاده قرار می‌گیرد:

- نخست، هر بردار داده، تنها شامل خوشه خودش است (یعنی برای هر بردار از داده‌ها یک خوشه داریم).
- انجام گام‌های بسیاری لازم است، در هر گام، دو خوشه نزدیک‌تر با یکدیگر ترکیب می‌شوند تا یک خوشهٔ مرکب را شکل دهند، در نتیجه دو خوشه به یک خوشه تبدیل می‌شود.
- هنگامی که تنها یک خوشه باقی ماند، همه نمونه‌ها را در ماتریس تمایز ترکیب می‌کنیم و متوقف می‌شویم.

نمونهٔ درخت تولید شده با این روش، در بخش بعدی خواهد آمد.

۲-۳. تحلیل خوشه‌ای داده‌های قرآنی

الگوریتم خوشه‌ای عامِ بالا بر یک نکتهٔ مهم سرپوش می‌گذارد: تعیین فاصله بین داده‌ها توسط جدول تمایز، اما فاصلهٔ میان خوشه‌های مرکب را روشن نمی‌کند و لازم است در هر گام، این فاصله محاسبه شود. چگونه این فاصله محاسبه می‌شود؟ این سؤال پاسخ ساده‌ای ندارد. تعاریف گوناگونی دربارهٔ ساختن خوشه وجود دارد و در هر کاربرد خاصی، مجازیم که یکی از آن تعاریف را انتخاب کنیم. مشکل این

است که ترکیب‌هایِ بی‌شمار معیار فاصله و تعاریف خوشه، سبب شده تا پژوهش گران، تحلیل‌های متفاوتی از داده‌های یکسان ارائه دهند. در حال حاضر، ملاک دقیقی برای انتخاب معیار فاصله و تعریف خوشه وجود ندارد. در حقیقت، این تعیین‌ناپذیری، مانع اصلی در استفاده از خوشه‌بندیِ سلسله مراتبی برای تحلیل داده‌ها می‌باشد. پژوهش حاضر، از کنار این نکتهٔ مهم عبور می‌کند؛ زیرا هدفش روش‌شناختی است: قصد نداریم در این مرحله از پژوهش، تحلیل خوشه‌ای قطعی از قرآن کریم ارائه دهیم، بلکه قصد داریم رویکرد به انجام این گونه مطالعات، را توسعه دهیم.[1] از این رو، یک ترکیب دقیق از معیار فاصله و تعریف خوشه به صورت تصادفی انتخاب شده و برای داده‌ها مورد استفاده قرار گرفته است که عبارتند از: مربع فاصلهٔ اقلیدسی و روش وارد.[2]

نتایج به شرح زیر می باشد (برچسب‌های A-D در سمت چپ، اشاره به مطلب بعدی دارند):

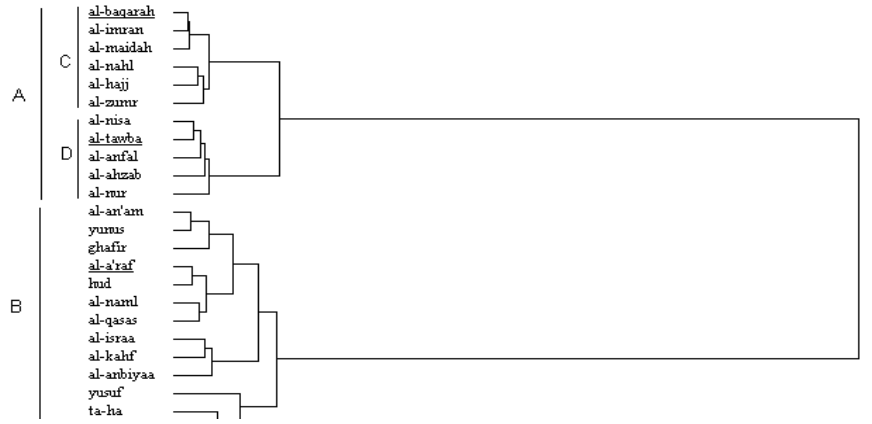
۱. روشن است که روی کرد به انجام این گونه مطالعات هنگامی توسعه پیدا می‌کند که نتایج آن قطعی و یا نزدیک به قطعی باشد، زیرا تحلیل‌های متفاوت از داده‌های یکسان، به جای آن‌که به مفسّر کمک کند و او را راهنمایی کند و سنگی را از سر راه مفسّر بردارد، مفسّر را در میان تحلیل‌های گوناگون سرگردان می‌سازد و بر سر راه او سنگ می‌نهد. البته ناگفته نماند که تحلیل‌های گوناگون هم‌چنان که مضراتی دارد، فایده‌ای هم دارد و آن این‌که این گونه تحلیل‌ها، ذهن مفسّر را نسبت به تفاسیر و برداشت‌های گوناگون از قرآن کریم نیز شکوفا می‌کند. (م)

نیز نویسنده این قسمت را بخوبی بیان نکرده و این ابهام بوجود آمده است که تغییر معیار تمایز (یا ماتریس تشابه) و تعریف خوشه می‌تواند نتیجه را تغییر دهد. البته این موضوع کاملا درست است. اما چون روش‌های آماری به صورت کلی و عمومی می‌باشند که در همه امور کاربرد داشته باشند، این جزء خصوصیات این روش‌ها است. لذا محقق بایستی با توجه به موضوع و هدف مطالعه بهترین ماتریس تشابه یا تمایز یا معیار فاصله را انتخاب نموده و همچنین خوشه را نیز با توجه به هدف و موضوع برگزیند. در این صورت هر روش دیگری مبتنی بر واقعیت نخواهد بود. (م)

2. Ward.

[DOR: 20.1001.1.17350689.1397.15.60.7.8]

[DOR: 20.1001.1.17350689.1397.15.60.7.8]



شکل۳: درخت ساخته شده بر اساس تحلیل خوشه‌ای

۳-۳. تفسیر

با فرض این که طول خطوط عمودی در بالای درخت، فاصلهٔ نسبی میان زیرخوشه‌ها را نشان می‌دهد، تفسیر این درخت بر حسب مناسبات سازه‌ایِ میان سوره‌ها آشکار است: دو زیرخوشهٔ اصلی A و B؛ A شامل دو زیر خوشهٔ C و D، و مانند آن زیر خوشه‌های دیگر است. هر چند شناخت ساختار سازه‌ایِ سوره‌ها، پیش‌شرط لازم فهمِ ارتباطات موضوعی میان سوره‌ها- هدفِ این کاربرد- است اما کافی نیست، زیرا این فهم، دربارهٔ ویژگی‌های موضوعی خوشه‌ها و تفاوت‌های موضوعی میان سوره‌ها و درون‌سوره‌ها، اطلاعاتی را به دست نمی‌دهد. این اطلاعات می‌تواند از معنی‌شناسیِ واژه‌ای برچسب‌های ستون‌ها در ماتریس داده‌ها به صورتی که در زیر آمده، استنتاج شود.

هر سطر در ماتریس داده‌ها، تصویر فراوانیِ واژه‌ای مترادف سوره است. از آن‌جا که تحلیل سلسله مراتبی، سطرهای ماتریس داده‌ها را بر حسب فاصلهٔ نسبی از یکدیگر در فضای داده‌ها خوشه‌بندی می‌کند، چنین برمی‌آید که در مجموعهٔ داده‌ها، تصاویر فراوانی واژه‌ای در خوشهٔ معینِ G ازهر تصویر دیگری، به یکدیگر نزدیکتر

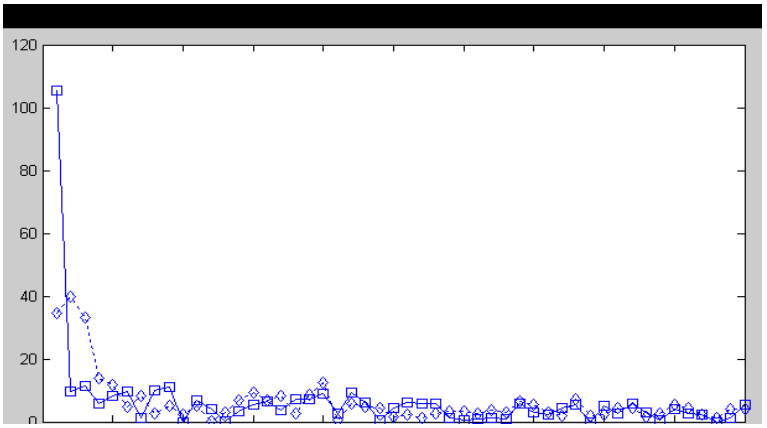
اند. تصاویر G می‌تواند به وسیلهٔ بُردار S خلاصه شود که ابعاد داده‌ها است و هر عنصرش، میانگین فراوانی را برای ستون ماتریس داده‌های مشابه در بر دارد:

$$S_j = (\sum_{i=1..n} F_{ij})/n$$

که در این جا j نشان دهنده j امین عنصر S، i نشان دهنده سطرهای ماتریس داده‌های F و n تعداد کل سطرهای خوشه‌ی G است. اگر S بر حسب معنی‌شناسی بر چسب‌های ستون ماتریس تفسیر شود، نمودار موضوعی‌ای برای G می‌شود: وابسته به تغییرات فراوانی s، فراوانی زیاد یک واژه نشان می‌دهد که این سوره‌هایی که G را تشکیل می‌دهند، به معنای صریح آن واژه پرداخته‌اند و نشانهٔ فراوانی کمِ یک واژه، نقیض آن است. این‌گونه نمودار موضوعی، می‌تواند برای هر زیر خوشه‌ای ساخته شود و تفاوت‌های موضوعی بین زیرخوشه‌ها می‌تواند از مقایسه نمودارها به دست آید.

بنابراین فرایند رایج برای تفسیر موضوعی درخت خوشه این است که ساختن و مقایسهٔ نمودارهای موضوعی برای زیرخوشه‌ها در هر سطح، از سطوح بالایِ درخت تا پایین درخت تا آن‌جا که مفید به نظر می‌آید، انجام شود.

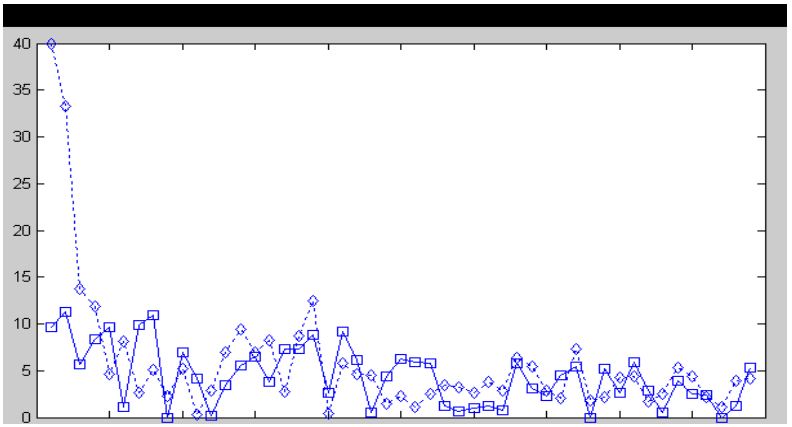
به عنوان مثال، کاربرد این فرایند رایج در زیر درخت A وB از خوشهٔ بالا را در نظر بگیرید. دو بردار میانگین فراوانی، یکی برای اجزاء سوره‌هایِ خوشهٔ A و یکی برای اجزاء سوره‌های خوشهٔ B ساخته شده است. سپس هر یک نسبت به دیگری روی نمودار نشان داده شده است؛ خطِ تو پر با شکل مربع، خوشهٔ A را نشان می‌دهد و خطِ نقطه‌چین با شکل لوزی، خوشهٔ B را نشان می‌دهد؛ برای روشنی، تنها پنجاه متغیر با بالاترین واریانس به ترتیب نزولی از چپ نشان داده شده‌اند:



شکل ۴: نمودار ابتدایی گروه‌های A و B

سوره‌های خوشهٔ A به صورت برجسته‌ای با معنای صریح متغیر۱ – متغیرِ با بالاترین تبیین واریانس در قرآن کریم- ارتباط بیشتری نسبت به سوره‌های خوشهٔ B دارند. این متغیّر، واژهٔ «الله» است که در قلب اسلام قراردارد؛ تفاوت در فراوانیِ وقوع واژهٔ «الله» در A و B، نخستین یافتهٔ معنادار روش‌شناسی حاضر است.

مقیاس‌بندی متغیّر «الله» بر همهٔ متغیّرهای دیگر تسلط دارد. برای افزایش دقت و وضوح سایر متغیّرها، «الله» از بردارهایِ فراوانیِ واژه‌ای حذف شد و بردارها دوباره بر روی نمودار نشان داده شدند:



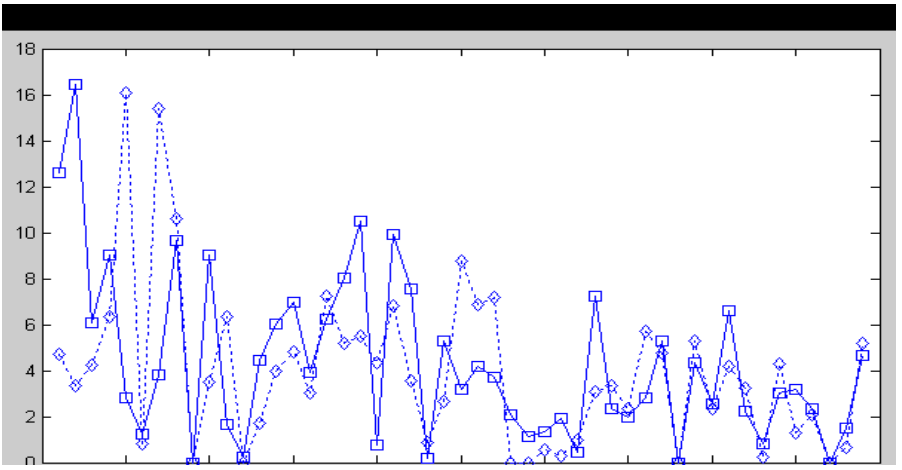
شکل۵: بازسازی نمودار گروه‌های A و B

اطلاع از زمینهٔ تاریخی از وحی قرآن کریم به حضرت محمدﷺ در این مرحله از تفسیر بسیار سخت است. سوره‌هایی که به حضرت محمدﷺ پیش از هجرتش به مدینه نازل شده است، سوره‌های «مکّی» نامیده می‌شوند و سوره‌های پس از هجرتش به مدینه، «مدنی» نامیده می‌شوند. سوره‌های مکی بر یگانگی و عظمت «الله»، وعدهٔ بهشت به درستکاران و هشدار عقاب به خطاکاران، تصدیق پیامبری حضرت محمدﷺ و آمدن رستاخیز، یادآوری پیامبران گذشته و حوادث زمان آن پیامبران به مردم تأکید دارند. از سوی دیگر، سوره‌های مدنی، جنبه‌های رسمی اسلام، وضع قوانین و آداب اخلاقی و معنوی، حقوق جنایی، اجتماعی، اقتصادی و سیاسی و رهنمودهای ارتباطات خارجی و مقررات نبردها و اسرای جنگی را به تصویر می‌کشند. این نتایج، از موارد مشخص شده در خوشه‌بندیِ ابتدایی در شکل ۳. مانند تفاوت موضوعی، به دست آمده است. همهٔ سوره‌ها در خوشهٔ A سوره‌های مدنی هستند (صرف نظر از سوره‌های «النحل» و «الزمر» که سوره‌های مکی هستند؛ در عین حال این سوره‌ها دارای آیاتی هستند که در مدینه نازل شده است). سیزده سوره‌ای که خوشهٔ B را تشکیل داده‌اند، همگی مدنی هستند. توزیع متغیّرها (کلیدواژه‌ها) در شکل ۵ بسیار معنادار است مانند متغیّر ۱ «قال» که در سوره‌های خوشهٔ B رایج است. سوره‌های این گروه، قصه‌های زیادی دارند که جنبه‌های مهمی از پیام قرآنی، یادکرد پیامبران گذشته و نبردهای‌شان و تحکیم رسالت اسلامی حضرت محمدﷺ را شرح می‌دهند. این، حاکی از کاربرد فعل «قال» به عنوان یک واژهٔ کلیدی در سبک روایت است. متغیر ۴ «قل» (گفتن دستوری) بیشترین تکرار را در گروه B نسبت به گروه A دارد. بیشترین نقل‌قول‌های سوره‌های مکی با واژهٔ «قل» آغاز می‌شود که دستوری برای حضرت محمدﷺ است تا با واژه‌های پس از این کلمه، شنوندگان‌اش را در موقعیت خاص مانند پاسخ به سؤالی که

[DOR: 20.1001.1.17350689.1397.15.60.7.8]

مطرح شده و یا یک حکم دربارهٔ ذات ایمان، مورد خطاب قرار دهد. کاربرد این واژه، با دعوت حضرت محمدﷺ به ایمان به خداوند و اسلام در سوره‌های مکی مناسب بود. متغیر ۵ «مؤمن»، متغیر ۸ «ایمان» و متغیر ۲۴ «تقوی»؛ در گروه A بسیار به کار رفته‌اند. سوره‌های مدنی، سوره‌هایی هستند که حضرت محمدﷺ در آن سوره‌ها، کسانی را مورد خطاب قرار می‌دهد که پیش از این به رسالت‌اش ایمان آورده بودند و حال وقت آن بود که به معرفی جنبه‌های اخلاقی و اجتماعی اسلام برای آنان بپردازد. دیگر متغیرهای رایج در گروه B متغیرهای ۱۴ و ۲۸ «آیات» و «آیه»اند. کاربرد این دو واژه برای حضرت محمدﷺ در دورهٔ اول اسلام در مکه خیلی مهم بود. او مجبور بود تا شاهد و نشانه‌ای به مردم ارائه دهد تا دعوت‌اش به ایمان به خدا و اسلام را تأیید کنند.

همان شیوهٔ خوشه‌بندی می‌تواند برای زیرخوشه‌های A و B به کار رود، دوباره مقیاس‌بندی «الله» که بر سایر متغیرها تسلط دارد، از میانگین بردارهای فراوانی حذف شود تا روشنی بهتری برای متغیرهای باقیمانده فراهم گردد. بردارهای فراوانی واژه‌ای برای C و D مثل حاصل زیر، بر روی نمودار نشان داده شود:



شکل۶: نمودار گروه‌های C و D

نتایج شکل ۶ پشتیبان ساختار موضوعی هر گروه هستند. سوره‌های گروه C بسیار فراوان از روایت و مخاطبْ ساختن حضرت محمدﷺ استفاده کرده‌اند تا زمینه را برای اثبات رسالتش به مردم، فراهم سازند. سوره‌های گروه D بیشتر مؤمنان را به خاطر پاداش اعمال صالحشان، مورد خطاب قرار می‌دهند. وقوع متغیرهای نسبی در این موضوعات، حاکی از چنین تفاوت‌هایی است.

۴. نتیجه و پیشنهادها

نتایج ابتداییِ بالا نشان می‌دهد که ترکیب و تفسیر معنایی درخت خوشه‌ای بر اساس فراوانی واژه‌ای، رویکرد مفیدی برای کشف موضوعی ارتباطاتِ میانْ سوره‌ایِ قرآن کریم است. اما این نتایج زمانی قابل استفاده‌اند که دو بحث اصلی حل شده باشد:

- استانداردسازیِ داده‌ها به خاطر اختلاف در طول سوره، همان‌گونه که در بخش ۳-۲ بحث شد.
- اختلاف در ساختار درخت با تفاوت ترکیب معیار فاصله و تعریف خوشه، همان‌گونه که در بخش ۲-۳ بحث شد.

پژوهش در موارد بالا، در حال پیشرفت است.

در پایان بیفزاییم که مشهور است که آنالیز خوشه‌ای سلسله مراتبی، به خاطر تفاوت معیار و تفاوت ترکیب قواعد خوشه‌ای، نتایج گوناگونی را به دست می‌دهد و در نتیجه نمی‌توان بر آن اعتماد کرد تا یک تحلیل قاطع ارائه دهد. گام بعدی این است که ببینیم آیا تحلیل مؤلفهٔ اصلی ماتریس فراوانی، نتایج سازگاری را با توجه به توصیفات بالا نتیجه می‌دهد. روش چند متغیرهٔ دیگری که برای این‌گونه داده‌ها به کار می‌رود، مقیاس‌بندی چند بعدی است. در دراز مدت، هدف این است تا با بهره‌گیری از روش‌های غیرخطی مانند نقشهٔ خودسازنده، محاسبات غیر خطی‌ای

[DOR: 20.1001.1.17350689.1397.15.60.7.8]

روی داده‌ها به دست دهد.

منابع

۱. بازرگان، مهدی (۱۳۵۵)، ***سیر تحول قرآن***، تهران: قلم.

۲. جرار، بسام نهاد (۱۴۱۴)، ***اعجاز الرقم ۱۹ تسعه عشر فی القرآن الکریم: مقدمات تنتظر النتائج***، بیروت: الموسسه الاسلامیه.

۳. جلالی نائینی، محمد رضا (۱۳۸۴)، ***تاریخ جمع قرآن کریم***، تهران: نشر سخن.

۴. دیدات، احمد (۱۳۸۱)، ***شبکه ریاضی قرآن: کشف سیستم حیرت انگیز ریاضی در قرآن***، ترجمه ساسان حبیب وند، قزوین: نشر طه.

۵. رفاعی، عدنان غازی (۱۴۲۱)، ***نظریه قرآنیه فی الاعجاز القرآنی: تعرض لاول مره فی العالم و نقد النظریه الاعجازیه فی القرآن الکریم***، دمشق: دارالفکر.

۶. روحانی، محمود (۱۳۶۸)، ***المعجم الاحصائی لالفاظ القرآن الکریم: فرهنگ آماری کلمات قرآن کریم***، مشهد: موسسه چاپ و انتشارات آستان قدس رضوی.

۷. سیوطی، جلال الدین (۲۰۰۰)، ***تناسق الدرر فی تناسب السور***، تحقیق السید الجمیلی، بیروت: مکتبه الهلال.

۸. صفوی، کورش (۱۳۷۹)، ***درآمدی بر معنی‌شناسی***، تهران: پژوهشگاه فرهنگ و هنر اسلامی/ انتشارات سوره مهر.

۹. عمر، احمد مختار (۱۳۸۵)، ***معناشناسی***، ترجمه سید حسین سیدی، مشهد: دانشگاه فردوسی مشهد.

۱۰. نوفل، عبدالرزاق (۱۳۸۳)، ***الاعجاز العددی فی القرآن الکریم: معجزه اعداد در قرآن کریم***، ترجمه سعید همایون، قم: بوستان کتاب قم.

۱۱. الهی‌زاده، محمد حسین (۱۳۹۱)، ***تدبّر میان سوره‌ای (لوح فشرده)***، مشهد: موسسه تدبر در قرآن و سیره.

۱۲. یزدانی، عباس (۱۳۷۵)، ***اعجاز عددی و نظم ریاضی قرآن***، کیهان اندیشه، شمارهٔ ۶۷، ۸۴– ۶۲.

1.Dror, J., Shaharabani, D., Talmon, R., Wintner, S.2004. *Morphological Analysis of the Qur'an*.

2 .Literary and Linguistic Computing, 19(4):431-452.

3 .Dunn, G. and Everitt, B. (2001).*Applied MultivariateData Analysis*, 2nd ed. Arnold, London.

4 .Everitt, B. (2001). *Cluster Analysis*, 4th ed. Arnold,London.

5 .Flynn, P., Jain, A., and Murty, M. (1999).*Dataclustering: A review*. In: ACM Computing Surveys31, 264–323.
6 .Gordon, A. (1999). *Classification*, 2nd ed. Chapman& Hall, London.
7 .Gore, P. (2000). *Cluster Analysis*. In H. E. A. Tinsley& S. D. Brown (Eds.), Handbook of applied
8 .multivariate statistics and mathematical modeling(pp. 297-321). Academic Press, San Diego, CA.
9 .Hair, H., Anderson, J., Black, W. and Tatham, R.(1998). *Multivariate Data Analysis*, 5th ed.Prentice-Hall International, London.
10 .Hand, D., Mannila, H., Smyth, P. (2001).*Principles of Data Mining*, MIT Press.
11.Kachigan, S .(1991) .*Multivariate Statistical Analysis .A conceptual introduction*. Radius Press, NewYork.
12 .Khalifa, Rashad,(1982), *Quran Visual Presentation of the miracle*, Arizona: Islamic Production.
13.Manning, C. and Schütze, H. (1999).*Foundations of Statistical Natural Language Processing*.Cambridge,Mass, MIT Press.
14.Oakes, M. (1998).*Statistics for Corpus Linguistics*.Edinburgh University Press, Edinburgh
15 .Thabet, N.(2004).*"Stemming the Qur'an*".In Proceedings of Arabic Script-Based Languages Workshop, COLING-04, Switzerland, August 2004.
16.Verleysen, M. (2003).*Learning high-dimensionaldata*. In: *Limitations and future trends in neuralcomputation*. IOS Press, Amesterdam, pp141-162.