

خوشه‌بندی مبتنی بر مدرک - اصطلاح: هم‌جواری موضوعات روان‌شناسی ازدواج در ادبیات

زیست پزشکی در دوره‌های زمانی «۱۹۹۹-۱۹۹۰» و «۲۰۰۸-۲۰۰۰»*

محمد توکلی‌زاده راوری^۱، مریم نجابتیان^۲

چکیده

مقدمه: روان‌شناسی ازدواج ریشه در مطالعات اجتماعی دارد. این موضوع در ادبیات زیست پزشکی نیز مورد توجه است. در این پژوهش سعی بر آن بود تا هم‌جواری و رابطه بین موضوعات روان‌شناسی ازدواج در ادبیات زیست پزشکی را بر اساس روش خوشه‌بندی مبتنی بر مدرک - اصطلاح تحلیل کند.

روش بررسی: سنجش هم‌جواری بین موضوعات حوزه‌ی روان‌شناسی ازدواج، بر اساس دو ماتریس هم‌جواری (Proximity matrix) صورت گرفته است. این ماتریس‌ها بر گرفته از محاسبه‌ی بسامد هم‌رخدادی (Co-occurrence) اصطلاحات موضوعی (Terms) و توصیفگرها (Descriptors) است که جزیی از دو مجموعه‌ی مدارک می‌باشد. این مجموعه‌ها به دوره‌های زمانی (۹۹-۱۹۹۰) و (۲۰۰۸-۲۰۰۰) تعلق داشت. مدارک موجود، حاصل نتایج دو جستجوی جداگانه در Medline بود. پس از تشکیل دو ماتریس ذکر شده، هم‌جواری موضوعات هر یک از دو مجموعه بر اساس روش کسینوسی سنجیده شد. خوشه‌بندی سلسله مراتبی (Hierarchical clustering) موضوعات هم‌جوار نیز با روش Ward صورت گرفت.

یافته‌ها: دو نمودار خوشه‌ای ماحصل سنجش هم‌جواری بین توصیفگرها بود. مقایسه‌ی این دو نمودار نشان داد که بیست توصیفگر از سی و هشت توصیفگر مورد مطالعه، پیوسته در هر دو دوره‌ی زمانی هم‌جواری خود را با یک یا چند توصیفگر دیگر حفظ کرده است. **نتیجه‌گیری:** می‌توان موضوعات روان‌شناسی ازدواج در زیست پزشکی را به پنج حوزه تقسیم کرد و با مقایسه‌ی دو نمودار خوشه‌ای حاصل شده، تغییر رابطه بین موضوعات هر حوزه با خودش و دیگر حوزه‌ها را بررسی کرد. نتایج این بررسی نشان داد که از روش خوشه‌بندی مدرک - اصطلاح می‌توان به برخی از سؤالات مرجع پاسخ گفت.

واژه‌های کلیدی: روان‌شناسی؛ ازدواج؛ طبقه‌بندی‌ها؛ پزشکی در ادبیات؛ مدلاین؛ سرعنوان‌های موضوعی.

نوع مقاله: تحقیقی

پدیرش مقاله: ۸۸/۶/۵

اصلاح نهایی: --

دریافت مقاله: ۸۷/۹/۲۰

ارجاع: توکلی‌زاده راوری محمد، نجابتیان مریم. خوشه‌بندی مبتنی بر مدرک - اصطلاح: هم‌جواری موضوعات روان‌شناسی ازدواج در ادبیات زیست پزشکی در دوره‌های زمانی "۱۹۹۰-۹۹" و "۲۰۰۸-۲۰۰۰". مدیریت اطلاعات سلامت ۳۸۹؛ ۷ (۲): ۱۸۶-۱۷۲.

مقدمه

زمینه‌ی ازدواج به دست جامعه‌شناسان صورت گرفته است که به مسایل جامعه‌شناختی (مانند درآمد، سن، دین، نژاد) و متغیرهای

تغییرات اقتصادی و اجتماعی قرن بیستم منجر به توجه عمومی نسبت به مسأله خانواده شده است. تمایل به فهم و حل مسایل مربوط به خانواده باعث افزایش مطالعات تجربی در این زمینه گردیده است (۱). از حدود دهه‌ی ۱۹۲۰ مطالعات مربوط به ازدواج آغاز گشت. توجه اولیه‌ی این مطالعات به رفتارهای جنسی بود (۲). به گفته‌ی Fincham و همکار، اغلب پژوهش‌های اولیه‌ی در

* این مقاله حاصل تحقیقی مستقل است که بدون حمایت مالی سازمانی انجام گرفته است.

۱. استادیار، علوم کتابداری و اطلاع‌رسانی، دانشگاه یزد، یزد، ایران. (نویسنده‌ی

مسئول). Email: tavakoli@vazduni.ac.ir

۲. کارشناس ارشد، روان‌شناسی، سازمان ملی جوانان، تهران، ایران.

روان‌شناختی (مانند شخصیت، احساس نسبت به درآمد، جامعه شناسی کودکی) می‌پرداختند که هنوز هم این گونه مطالعات ادامه دارد (۳). توجه به مسایل رفتاری، محرکی برای روان‌شناسان بود تا به مبحث ازدواج به طور خاص بپردازند. عامل دیگر پرداختن روان‌شناسان به مسأله ازدواج، تحقیقات گروهی در روان‌شناسی اجتماعی بود که با به کارگیری تئوری سیستم‌های عمومی (General systems theory) به مطالعه‌ی خانواده می‌پرداخت.

گرچه روان‌شناسی ازدواج ریشه در تحقیقات جامعه‌شناسان و روان‌شناسان اجتماعی دارد، اما امروزه ادبیات مربوط به زیست پزشکی نیز به طور گسترده به این موضوع پرداخته است، به ویژه در مواردی که مربوط به مسأله شناخت و درمان مسایل روانی ازدواج می‌باشد. از این رو این پژوهش بر آن است تا موضوعات مطرح روان‌شناسی ازدواج در ادبیات زیست پزشکی را بر اساس اصطلاحات موضوعی برگرفته از متون مربوط به این زمینه دسته‌بندی و تحلیل کند تا به توان سیر تحول روابط موضوعی در این زمینه را کشف کرد. این تحلیل بر مبنای مقایسه‌ی دسته‌های موضوعی حاصل شده از دو دوره‌ی زمانی (سال‌های ۹۹-۱۹۹۰ و ۲۰۰۸-۲۰۰۰) انجام گرفته است.

دسته بندی موضوعات این رشته بر پایه‌ی فنون خوشه‌بندی خودکار (بدون توجه به دسته‌بندی‌های نظری) انجام شده است. زیرا ابزار و روش‌های کشف دانش از متون نوشتاری مانند خوشه‌بندی، به محققین کمک می‌کند تا اطلاعات را مدیریت کنند و واقعیت‌ها و ارتباطاتی را در ادبیات زیست پزشکی کشف کنند که به حل مسایل زیستی کمک می‌کند (۴).

جهت انجام این کار از روش سنجش هم‌جواری بین اصطلاحات موضوعی مستخرج از متن مدارک (به ویژه مقالات) بهره‌جویی شده است که در تحقیقات مربوط به اطلاع‌رسانی و کتاب‌سنجی کاربرد فراوان دارد. از این رو ابتدا به معرفی مسایل و موضوعات پیرامون این فنون پرداخته شده است.

این نوشتار به گونه‌ای تنظیم گردیده است که هم به متخصصان و پژوهشگران حیطه روان‌شناختی کمک کند و هم

واقعیتی را به متخصصان امر کتابداری و اطلاع‌رسانی نشان دهد که چگونه می‌توان از این روش در امور مربوط به این رشته، به ویژه کار مرجع بهره جست. چرا که معرفی برخی از موضوعات به کار رفته در ادبیات خوشه‌بندی، می‌تواند راهی را برای انجام پژوهش‌های بعدی به متخصصان این دو رشته عرضه کند.

خوشه‌بندی مبتنی بر مدرک - اصطلاح:

خوشه‌بندی همان دسته‌بندی اشیا است و متغیرها بر پایه میزان شباهت یا فاصله ویژگی‌های مورد نظر آن‌ها استوار می‌باشد. از طرفی، تحلیل خوشه‌ای نوعی طبقه‌بندی است که از طریق تحلیل مجموعه‌ای از داده‌های پیچیده، گروه‌هایی متجانس ایجاد می‌کند (۵). به فرض اگر هدف دسته‌بندی افراد را بر اساس بهره هوشی آن‌ها باشد، ویژگی مورد نظر هوش بهره آن‌ها است. لذا هر چه بهره هوشی این افراد به هم نزدیک‌تر باشد در دسته‌بندی از لحاظ مکانی نزدیک به یکدیگر قرار می‌گیرند، یعنی هر چه تفاوت این مقدار بین دو فرد کمتر باشد، شباهت آن‌ها بر اساس این ویژگی بیشتر یا به عبارتی عدم شباهت (فاصله) آن‌ها کمتر است. در خوشه‌بندی خودکار، هر عضو از یک مجموعه مورد تحقیق یک بردار (Vector) هندسی در نظر گرفته می‌شود که نقاط تشکیل دهنده‌ی این بردار میزان اختلاف بین آن عضو با هر یک از اعضای دیگر آن مجموعه است. هر چه دو بردار نقاط مشترک بیشتری داشته باشند، آن دو به هم شبیه‌ترند یا فاصله (Distance) آن دو کمتر است. از این جهت می‌توان گفت که خوشه‌بندی، مقایسه‌ی شباهت (Similarity) یا به عبارتی فاصله هر بردار با دیگر بردارها می‌باشد (۶). بنابراین، وقتی که گفته می‌شود دو اصطلاح شبیه هستند، بدان معنی نیست که آن‌ها مترادفند یا از لحاظ معنایی شبیه‌اند (۷) و یا لزوماً رابطه عام و خاص بین آن‌ها وجود دارد، بلکه از لحاظ ادبیات آماری بدان معنی است که میزان نقاط مشترک آن‌ها روی بردار چه اندازه است.

برای خوشه‌بندی و طبقه‌بندی متون مربوط به یک زمینه‌ی موضوعی خاص، هر کدام از اصطلاحات موضوعی موجود در آن‌ها یک بردار در نظر گرفته می‌شود (۸)، سپس تعیین ارتباط بین اصطلاحات گوناگون بر پایه‌ی میزان هم‌رخدادی اصطلاحات یا

طریق اصطلاحات برجسته و مشترک درون آن‌ها خوشه‌بندی می‌شوند (۴).

مزایای خوشه‌بندی سلسله‌مراتبی مبتنی بر مدرک - اصطلاح: خوشه‌بندی به دو صورت سلسله‌مراتبی و غیر سلسله‌مراتبی امکان‌پذیر است. روش سلسله‌مراتبی هنگامی مؤثر است که تعداد موردها کم باشد. در معمول تعداد کمتر از ۱۰۰۰ را توصیه می‌کنند. خوشه‌بندی غیر سلسله‌مراتبی وقتی ارزشمند است که بر اساس دانش قبلی به توان تعداد خوشه‌ها را به احتمال تعیین کرد. به طور معمول این روش برای زمینه‌هایی مثل آزمودن صحت یک طبقه‌بندی مناسب است (۵). همچنین در زمانی که تعداد موارد مورد بررسی زیاد است این نوع خوشه‌بندی توصیه می‌شود.

Tao و همکاران، مزیت خوشه‌بندی سلسله‌مراتبی بر اساس روش‌های آماری و خودکار را در برابر روش‌های متخصص محور (دستی) چنین برمی‌شمارند: ۱. خوشه‌بندی بر اساس روش‌های دستی بسیار هزینه‌بر است، زیرا این فرآیند نیازمند متخصصین موضوعی است تا ارتباط موضوعات و مدارک را بیانند. ۲. ممکن است گستره‌ای از مدارک وجود داشته باشد که به صورت طبیعی نتوان آن‌ها را در هیچ سلسله‌مراتب موضوعی جای داد و حتی امکان دارد متخصصین موضوعی در ارزیابی ارتباط معنایی بین آن‌ها دچار مشکل شوند. ۳. توسعه‌ی خودکار خوشه‌های سلسله‌مراتبی بر پایه‌ی آمارهایی است که از مدارک برگرفته شده است و می‌تواند به صورت آماری ارتباط بین سلسله‌خوشه‌ها را نشان دهد. آن‌ها همچنین بر پایه‌ی دو دلیل خوشه‌بندی سلسله‌مراتبی را بسیار سودمند می‌دانند: ۱. بسیاری از کاربران اطلاعات ترجیح می‌دهند تا از طریق بررسی فهرست‌های سلسله‌مراتبی به دنبال اطلاعات مورد نیاز خود بگردند. ۲. ساختار سلسله‌مراتبی، ارتباط بین طبقه‌های مرتبط را مشخص می‌سازد و اطلاعات ارزشمندی را برای حل بسیاری از مسایل به دست می‌دهد (۱۱).

از جهتی نیز با رشد فراوان اطلاعات به واسطه‌ی حضور

واژه‌های موجود در یک متن یا مجموعه‌ای از متون است که بدون تکیه به روش‌های پیچیده‌ی منبعث از تئوری‌های زبان‌شناسی صورت می‌گیرد (۹).

نتیجه‌ی محاسبه‌ی این هم‌رخدادی در یک ماتریس مربع، معروف به ماتریس هم‌جواری، نشان داده شده است.

جدول ۱: نمونه‌ای از یک ماتریس هم‌جواری

	Affect	Anger	Anxiety	...
Affect	۷۹	۲	۴	...
Anger	۲	۱۹	۴	...
Anxiety	۴	۴	۵۶	...
...

جدول ۱ بریده‌ای از یک ماتریس هم‌جواری است. نقطه‌ی تلاقی دو ستون افقی و عمودی، بسامد مدارکی (بسامد هم‌رخدادی) است که دو اصطلاح مورد نظر هم‌زمان در آن‌ها وجود دارد. به طور مثال اگر در جدول ۱ از ردیف سوم Anger و از ستون دوم Affect انتخاب شود، مشاهده می‌شود که ۲ مدرک در آن مجموعه وجود داشته است که هم‌زمان حاوی این دو اصطلاح بوده‌اند.

خوشه‌بندی بر اساس ماتریس مبتنی بر مدرک - اصطلاح (مانند جدول ۱) دو حالت دارد: ۱. خوشه‌بندی مدارک و ۲. خوشه‌بندی اصطلاحات. ساده‌ترین روش برای خوشه‌بندی، خوشه‌بندی مدارک است. در این حالت، مدارک بر اساس شباهت بسامدی اصطلاحاتشان در خوشه‌های متمایز گروه‌بندی می‌شود. برای این کار اصطلاحات به طور معمول به عنوان مورد یا برچسب در نظر گرفته می‌شوند. روش ساده‌ی دیگر، خوشه‌بندی اصطلاحات است. اصطلاحات بر اساس شباهتی که بین مدارک منتسب به آن‌ها وجود دارد، در خوشه‌های متمایز گروه‌بندی می‌شود، سپس مدارک مربوط به آن‌ها در خوشه‌هایی که آن اصطلاحات وجود دارد، قرار می‌گیرد (۱۰). یعنی مدارکی که محتوای مشترکی دارند از

علمی یا یک تئوری روشی خاص برای طبقه‌بندی موضوعات شاخه‌ای از روان‌شناسی پیشنهاد می‌کند، از روش خوشه‌بندی می‌توان آن طبقه‌بندی‌ها را آزمود. ۳- ساده‌سازی. خوشه‌بندی یک روش قدرتمند برای ساده‌سازی مجموعه‌ای پیچیده از داده‌ها است. از آن می‌توان در زمانی که قضاوت ذهنی انسانی برای گروه‌بندی اشیاء دشوار و یا ناممکن است، بهره برد (۵).

Stephen به کاربرد خوشه‌بندی در ادبیات مربوط به ازدواج و خانواده اشاره دارد. او در اثر خود سعی دارد تا بر اساس این الگو رابطه‌ی بین مفاهیم تئوریک در حوزه‌ی مورد تحقیق خود را تعیین و معرفی کند (۱۵). از این رو می‌توان یکی دیگر از کارکردهای خوشه‌بندی را در روان‌شناسی و ازدواج تعیین کرد و معرف رابطه‌ی بین مفاهیم تئوریک این موضوعات دانست. علاوه بر آن، با خوشه‌بندی می‌توان اصطلاحات موضوعی درون مدارک در چند دوره‌ی مختلف را به صورت جداگانه مقایسه و خوشه‌بندی کرد. با استفاده از دسته‌بندی‌های حاصل از آن خوشه‌ها می‌توان به تحلیل ثبات یا تغییر رابطه‌ی بین موضوعات مختلف در طول زمان پرداخت و نحوه‌ی تعامل موضوعات مختلف را با یکدیگر کشف کرد. این پژوهش بر این جنبه از کارکرد خوشه‌بندی تأکید دارد.

یکی از معایب اصلی خوشه‌بندی، مشکل ایجاد خوشه‌های معنی‌دار و قابل فهم است که نمی‌توان به راحتی آن‌ها را نام‌گذاری کرد یا بر چسب یا عنوانی مناسب به آن‌ها نسبت داد (۱۰). نکته دیگر این است که به طور معمول مدارک، دارای بیش از یک موضوع هستند، از این جهت یک مدرک باید هم‌زمان در چند خوشه قرار بگیرد (۱۰). اما به اجبار تنها در یک خوشه قرار می‌گیرد. خوشه‌بندی بر اساس روش فازی می‌تواند این هم‌پوشانی را به تصویر بکشد (۱۶)، اما این روش پیچیده و فاقد سایر مزیت‌های خوشه‌بندی سلسله‌مراتبی است.

روش‌های سنجش (Measurement) هم‌جواری و خوشه‌بندی:

معمول‌ترین روش سلسله‌مراتبی، محاسبه‌ی تراکمی است. در این روش ابتدا موردهای بسیار نزدیک به هم مشخص می‌شود و قدم به قدم هر لایه (Layout) با لایه‌ای دورتر تلفیق

اینترنت، مرزهای موضوعی در هم تنیده شده است. برای حل این مسأله Koller و Sahami (۱۲) استفاده از ساختار خوشه‌بندی سلسله‌مراتبی را توصیه می‌کنند. همه این‌ها دلیلی است بر این که چرا «خوشه‌بندی به سرعت در بسیاری از زمینه‌ها از جمله علوم اجتماعی رشد یافته است» (۵).

یکی از کارکردهای خوشه‌بندی سلسله‌مراتبی مبتنی بر مدرک-اصطلاح، در جایی است که روش‌های کتاب‌سنجی و علم‌سنجی قادر به پاسخگویی نیست. به طور مثال از طریق خوشه‌بندی می‌توان ارتباطات پنهان بین اصطلاحات موجود در مجموعه‌ای از مدارک را یافت که به طور مستقیم یا از طریق آن روش‌ها قابل دیدن نیستند. به عنوان نمونه، Swanson و همکار، از این طریق ارتباط بین بیماری Raynaud و اسیدهای چرب موجود در روغن ماهی را پیدا کردند (۱۳). ارزش کار آن‌ها در این بود که تا آن زمان در هیچ مدرک یا جمله‌ای، ارتباطی بین آن دو ذکر نشده بود (۴).

اگر مدارک بر اساس هم‌پوشانی استنادات خود شباهت سنجی شوند، ممکن است میزان شباهت صفر نشان داده شود. ولی در مقابل از طریق تحلیل ارتباط موضوعی، بر پایه‌ی اصطلاحات موجود در متن، می‌توان شباهت فراوانی بین دو مدرک یافت (۱۴). از طرفی تحلیل‌های علم‌سنجی که به تأثیرگذاری نشریات، نویسندگان و مقالات می‌پردازد، فاقد این توانایی هستند که نشان دهند چه مفهوم، چه موضوع و چه نظریه‌ای در کجا و در چه زمان منتشر شده است (۱۵).

در حوزه‌های روان‌شناسی و ازدواج نیز به مبحث خوشه‌بندی پرداخته شده است. این روش در این حوزه‌ها کارکردهای گوناگونی دارد. Borgen و همکار، سه کارکرد آن را مورد توجه قرار می‌دهند: ۱. کارکرد توضیحی. اگر زمینه‌ی مورد تحقیق به نسبت جدید باشد، خوشه‌بندی قدم ابتدایی برای تشخیص می‌باشد و ساختار بندی گروه‌های فرعی، موضوعاتی است که به طور بالقوه در فهم مسایل تحقیق حول آن محور ارزشمندند. به عبارتی، محققین از خوشه‌بندی برای خلاصه‌سازی محتوای یک مجموعه‌ی بزرگ از داده‌ها بهره می‌برند، ۲. آزمودن طبقه‌بندی‌های قبلی. اگر دستاوردهای

در شکل ۱، طول خطوط موازی که با یک خط عمودی به هم مرتبط شده‌اند، میزان شباهت یا به عبارتی فاصله‌ی مجاورتی آن‌ها را ترسیم می‌کند. به طور مثال از این طریق می‌توان فهمید که شباهت پروانه با پیمانه، نسبت به پوریا و پگاه بیشتر و فاصله‌ی آن‌ها کمتر است.

روش ترکیب خوشه‌ها با یکدیگر نیز باید در خوشه‌بندی لحاظ شود. آن چه که در مثال بالا آمد، روشی ساده برای ترکیب خوشه‌های یک مسأله‌ی بسیار ساده بود. باید توجه داشت که برای این کار روش‌های متعددی وجود دارد که هر کدام ویژگی منحصر به خود را دارد. پرکاربردترین آن‌ها روش‌های Ward، Ward نزدیک‌ترین هم‌جواری، Ward دورترین هم‌جواری و Ward است. در روش اول برای ترکیب خوشه‌ها، فاصله‌ی دو بردار، فاصله‌ی بین دو نقطه‌ای از آن دو بردار است که کمترین فاصله را با هم دارند. بر خلاف آن در روش دورترین هم‌جواری، تعیین فاصله‌ی بین دو بردار بر اساس دو نقطه‌ای صورت می‌گیرد که بیشترین فاصله را از یکدیگر دارند. روش Ward، مشهور به ارایه دهنده آن Ward (۱۹۶۳)، عبارت است از جمع مربع فاصله‌ی اقلیدسی هر مورد در هر خوشه با میانگین همه‌ی متغیرها (۱۷). به عبارتی این روش فاصله‌ی بین دو خوشه را بر اساس فاصله‌ی بین عناصر موجود در آن خوشه‌ها در نظر می‌گیرد (۶). روش Ward باعث می‌شود تا پراکندگی درون خوشه‌ها در هر مرحله از گروه‌بندی کمتر (۱۸، ۵) و خوشه‌های ریزتر با حجم کم ایجاد شود. بنابراین، در جایی که تعداد موارد کم است و به خوشه‌بندی بسیار ریز نیاز است این روش می‌تواند مؤثر باشد.

نکته‌ی مهم دیگر در خوشه‌بندی، انتخاب روشی مناسب برای سنجش فاصله‌ی بین بردارها است، برای این کار روش‌های متفاوتی موجود است که هر یک ویژگی منحصر به خود را دارد (۵). مشهورترین این روش‌ها عبارت است از فاصله‌ی اقلیدسی (Euclidean distance) و شباهت کوسینوسی (Cosine similarity). معمول‌ترین این روش‌ها به ویژه در سنجش فاصله‌ی بین بردارها در موضوعات مربوط به کتاب‌سنجی و علم‌سنجی محاسبه‌ی فاصله اقلیدسی است.

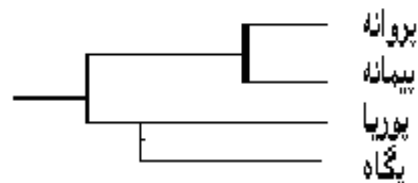
می‌شود. بنابراین، اگر n مورد موجود باشد، فرآیند تلفیق با در کنار هم گذاشتن دو موردی که بیشترین نزدیکی را به یکدیگر دارند شروع می‌شود و اولین خوشه را شکل می‌دهند. در مرحله‌ی بعدی، موردی که در یک لایه‌ی دورتر با آن‌ها ارتباط دارد، به آن‌ها وصل می‌شود و این عمل $n-1$ بار صورت می‌گیرد تا خوشه‌بندی کامل شود (۵). برای روشن‌تر شدن موضوع به یک مثال پرداخته می‌شود. اگر چهار نفر را بر اساس هوش‌بهرشان به صورت سلسله مراتبی خوشه‌بندی شوند. جدول ۲ هوش‌بهر هر کدام از آن‌ها را نمایش می‌دهد:

جدول ۲: هوش‌بهر فرضی ۴ نفر

نام	هوش‌بهر
پروانه	۲۰۰
پگاه	۱۰۰
پیمانه	۲۰۰
پوریا	۱۲۰

در مرحله‌ی اول مشاهده می‌شود که پروانه و پیمانه کمترین اختلاف را از این حیث دارند (اختلاف صفر). بنابراین، این دو در یک خوشه قرار می‌گیرند (خوشه ۱). در مرحله‌ی بعدی کمترین اختلاف بین پوریا و پگاه است (اختلاف ۲۰). این دو نیز در یک خوشه اما با فاصله‌ی بیشتر قرار می‌گیرند (خوشه ۲).

در لایه‌ی بعدی از طریق پوریا در خوشه‌ی دو با پروانه و پیمانه در خوشه‌ی یک، خوشه‌ای وسیع‌تر ایجاد می‌شود. این خوشه‌بندی سلسله مراتبی ساده که بر اساس محاسبه‌ی تراکمی صورت گرفته است که در شکل ۱ به نمایش در آمده است.



شکل ۱: نمونه‌ای از یک خوشه‌بندی سلسله مراتبی بر اساس روش محاسبه‌ی تراکمی

مورد، اصطلاح کنترل شده‌ی Marriage به صورت Marriage [MeSH Headings] جستجو شد. اصطلاح ذکر شده درون قلاب، دستوری است که از سیستم می‌خواهد تا تنها مدارکی را بازبایی کند که در جز توصیفگر آن‌ها موضوع یاد شده یافت می‌شود. به زبان دیگر، جستجو را در جز سرعنوان‌های موضوعی پزشکی (MeSH Headings) محدود می‌کند. تفاوت این دو جستجو محدود کردن تاریخ انتشار مدارک در دو دوره‌ی متفاوت بود. ابتدا مدارک منتشره بین سال‌های ۹۹-۱۹۹۰ و سپس مدارک منتشره بین سال‌های ۲۰۰۸-۲۰۰۰ بازبایی شده است. نتیجه‌ی جستجوی اول ۳۳۵۷ مدرک و جستجوی دوم ۲۵۱۰ مدرک بود (۱). هر یک از این نتایج به صورت جداگانه در دو فایل متنی ذخیره گردید.

در مرحله‌ی بعدی باید در خصوص نحوه‌ی استخراج اصطلاحات موضوعی از متون بازبایی شده تصمیم‌گیری به عمل می‌آمد.

دو روش مهم برای استخراج موضوعات مهم یک مدرک وجود دارد: ۱- روش خودکار مبتنی بر نمایه‌سازی ماشینی، که در معمول بر پایه‌ی بسامد واژه‌ها، اصطلاحات موجود یک متن (عنوان، چکیده، ...) و وزن‌دهی (Weighting) به آن‌ها صورت می‌گیرد. مشهورترین آن‌ها، بهره‌گیری از الگویی است که مبتنی بر وزن سراسری و محلی است. از این طریق وزن واژه‌ها و اصطلاحات بر پایه‌ی فرمول معروف TF-IDF محاسبه می‌شود. ۲- روش دیگر استفاده از توصیفگرهایی است که به صورت کنترل شده توسط نمایه‌سازان از طریق واژه‌های کنترل شده به مدارک داده می‌شود.

تصمیم‌گیری در خصوص سودمندی و بهتر بودن هر یک از این روش‌ها بحث انگیز است. به طور مثال Buscaldi و همکاران، برای خوشه‌بندی موضوعات، استخراج کلید واژه‌ها را توصیه می‌کنند؛ چرا که معتقدند که اگر در یک زمینه موضوعی بسیار خاص عمل خوشه‌بندی انجام شود، نتایج غیر دقیق و ناستواری به دست می‌دهد (۸). در مقابل Stephen معتقد است که هر چه زبان مجموعه‌ی متون مورد بررسی کنترل شده‌تر باشد بهتر است (۱۵).

در این روش از یک جفت بردار مورد سنجش، یکی به عنوان متغیر X و دیگری به عنوان متغیر Y در نظر گرفته می‌شود، سپس محاسبه‌ی فاصله‌ی اقلیدسی بین دو مورد بر اساس جذر مجموع مربع X ها منهای Y ها صورت می‌گیرد. در این روش سعی بر آن است تا میزان عدم شباهت (فاصله) بین دو مورد محاسبه شود. بنابراین، ضرایب حاصل هر چه بزرگتر باشد، فاصله بیشتر و شباهت کمتر است:

فرمول ۱: محاسبه‌ی فاصله‌ی اقلیدسی

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

روش کوسینوسی نیز یکی از روش‌های مشهور دیگر است که به طور گسترده در مبحث بازبایی اطلاعات از آن استفاده می‌شود (۶). در این روش نیز یک جفت بردار مورد سنجش یکی به عنوان متغیر X و دیگری به عنوان متغیر Y در نظر گرفته می‌شود و مشابهت بین دو بردار بر اساس محاسبه‌ی کوسینوس زاویه‌ی حاصل از دو بردار صورت می‌گیرد، این روش بر خلاف فاصله‌ی اقلیدسی بر شباهت بین دو بردار تکیه دارد. بنابراین، ضرایب حاصل هر چه بزرگتر باشد، فاصله کمتر و شباهت بیشتر است:

فرمول ۲: محاسبه‌ی شباهت کوسینوسی

$$\text{sim}(x, y) = \frac{\sum_i (x_i y_i)}{\sqrt{(\sum_i x_i^2)(\sum_i y_i^2)}}$$

روش بررسی

این پژوهش توصیفی از نوع کاربردی-مقطعی بود. Medline یکی از مشهورترین پایگاه‌های اطلاعاتی در زمینه‌ی زیست پزشکی است که تا به امروز بیش از هفده میلیون مدرک نمایه شده را در خود جای داده و از این رو تبدیل به یک پایگاه جامع و پر کاربرد شده است. در این تحقیق از آن به عنوان منبع استخراج داده‌های مورد نیاز استفاده شد. برای این کار دو جستجوی جداگانه از طریق PubMed صورت گرفت. در هر دو

بر اساس داده‌های مستخرج از مجموعه‌ی مدارک ۹۹-۱۹۹۰ و جدول دوم از داده‌های مستخرج از مجموعه‌ی مدارک ۲۰۰۸-۲۰۰۰ پر شد. قابلیت جستجوی خودکار با دستورات SQL query search، در این جدول‌ها، امکان تهیه‌ی دو جدول مربع شکل یا دو ماتریس هم‌جواری را برای دو دوره‌ی یاد شده فراهم کرد. برای تهیه‌ی چنین ماتریسی‌هایی، بسامد هم‌رخدادی هر توصیفگر با ۳۷ توصیفگر دیگر سنجیده شد. از این جهت ایجاد هر ماتریس حاصل 38×38 بار جستجوی جفت به جفت در جداولی بود که در SQL server ایجاد شده بود. یعنی محاسبه‌ی بسامد هم‌رخدادی توصیفگر اول با توصیفگر اول تا سی و هشتم، توصیفگر دوم با توصیفگر اول تا سی و هشتم و الی آخر.

اساس کار محاسبات این پژوهش، دو ماتریس مشابه جدول ۱ بود که یکی به دوره‌ی ۹۹-۱۹۹۰ و دیگری به ۲۰۰۸-۲۰۰۰ تعلق داشت. خوشه‌بندی ۳۸ موضوع موجود در هر یک از این ماتریس‌ها از طریق برنامه‌ی آماری SPSS به روش سلسله مراتبی صورت گرفت. برای این کار، هر مورد (توصیفگر) در هر ردیف به عنوان یک بردار و بسامد مدارک در سلول‌های روبه‌رویی، متغیرهای آن بردار محسوب شد.

روش سنجش و خوشه‌بندی در این پژوهش:

پس از بررسی روش‌های گوناگون، مشخص گردید که سنجش فاصله‌ی بین بردارها از روش کوسینوسی، به صورت بسیار منطقی‌تری توصیفگرهای مربوط به موضوع مورد تحقیق را دسته‌بندی می‌کند. به عبارتی دسته‌بندی موضوعی از این روش با دسته‌بندی ذهنی مطابقت بیشتری داشت. به همین جهت محاسبه‌ی کوسینوس زاویه‌ی بردارها برای سنجش فاصله‌ی بین آن‌ها مناسب تشخیص داده شد.

از آن‌جا که بررسی توصیفگرهای مناسب با موضوع روان‌شناسی ازدواج در ادبیات زیست پزشکی منجر به انتخاب ۳۸ توصیفگر شده بود، تصمیم گرفته شد تا برای ترکیب خوشه‌ها از روشی استفاده شود تا خوشه‌های کوچکتر و معنی‌دارتری به دست دهد. روش Ward برای ترکیب خوشه‌ها در این تحقیق مناسب‌تر تشخیص داده شد.

نمایه‌سازی مدارک در پایگاه Medline بسیار دقیق انجام می‌گیرد و عمق نمایه‌سازی (In-depth indexing) در آن در حد قابل قبولی است (به طور متوسط دوازده توصیفگر در ازای هر مدرک). بنابراین، جامعیت (Exhausticificity) و مانعیت (Specificity) نمایه‌سازی در آن قابل توجه است و در حد لازم جنبه‌های مختلف موضوعات یک مدرک را پوشش می‌دهد. از این جهت، متوسط تعداد توصیفگرهای داده شده به یک مدرک در Medline به پژوهشگران این امکان را می‌دهد تا با اطمینان از آن‌ها به جای کلیدواژه‌های آزاد در این پژوهش همانند Mc Hehen berger و سایرین بهره گرفت.

پس از تصمیم‌گیری درباره‌ی نحوه‌ی استخراج موضوعات از متون، از طریق یک برنامه‌ی کامپیوتری واسط، توصیفگرهای هر یک از مدارک تک به تک استخراج گردید. هدف از این کار تهیه‌ی یک واژگان از موضوعاتی بود که در ادبیات ازدواج به کار رفته است. از به هم آمیخته کردن واژه‌های برگرفته از مجموعه‌ی مدارک دو دوره یاد شده، واژگانی متشکل از حدود ۸۵۰ توصیفگر متمایز به وجود آمد. پس از آن، با نظر یک روان‌شناس ارشد و مشغول در موضوع ازدواج، موضوعات خاص روان‌شناسی در ادبیات روان‌شناسی ازدواج از میان آن‌ها استخراج گردید. سپس، این توصیفگرها در سرعنوان‌های موضوعی پزشکی (MeSH) کنترل شد تا یقین حاصل شود که آیا مفهوم هر یک از توصیفگرهای انتخاب شده همان است که متخصص موضوعی مد نظر دارد. در کنار آن سعی شد تا حد ممکن در این لیست افزونگی (Redundancy) وجود نداشته باشد. به طور مثال بین توصیفگر Anger و Emotion رابطه‌ی عام و خاص وجود دارد که از این میان Emotion حذف شد. سرانجام ۳۸ توصیفگر برای تحقیق برگزیده شد.

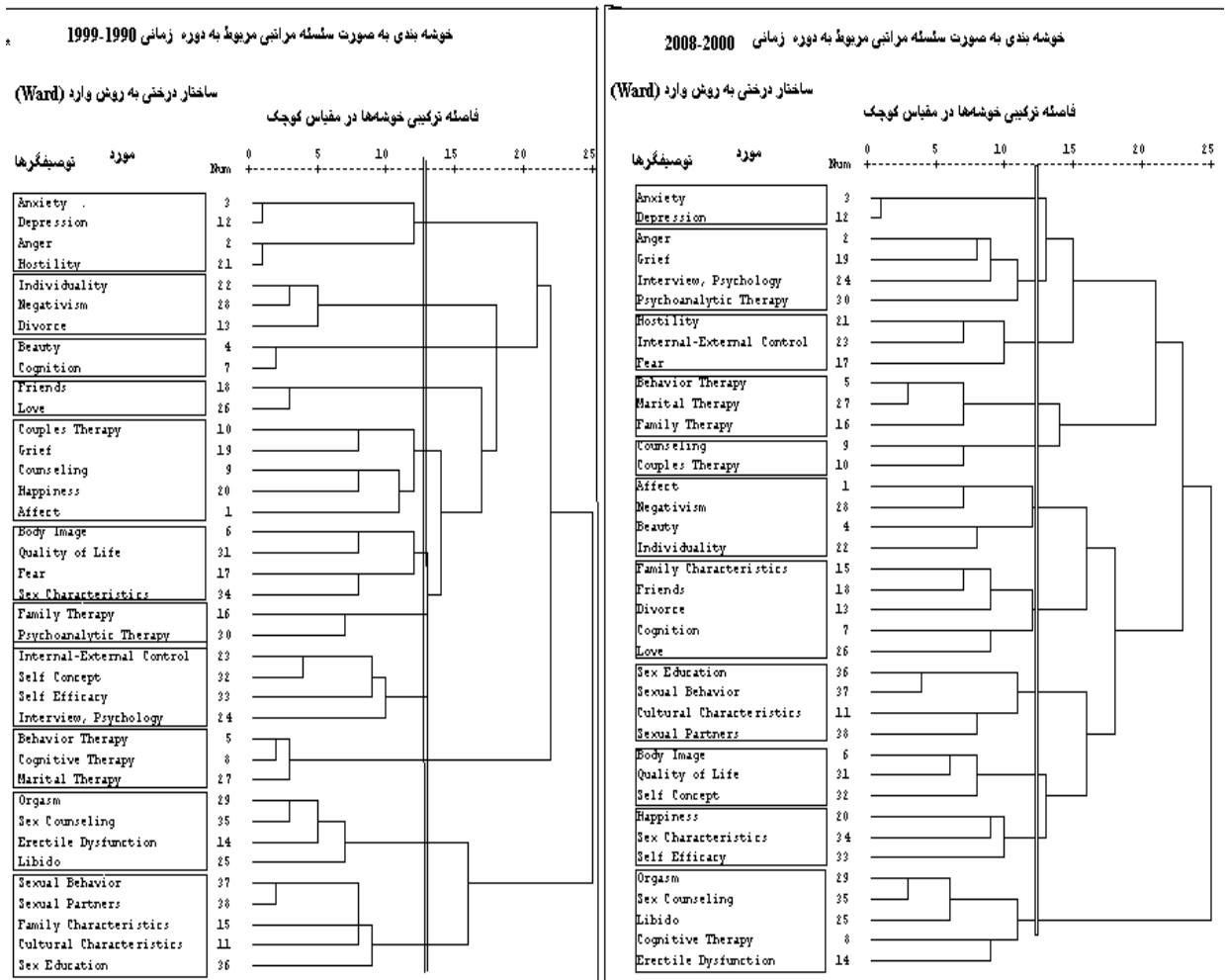
سپس پایگاهی متشکل از دو جدول در SQL server ایجاد شد. هر یک از جداول دارای دو ستون بود: ۱. شماره‌ی ردیف مدرک در فایل کامپیوتری مربوط و ۲. توصیفگر. از طریق یک برنامه‌ی کامپیوتری دیگر، ستون‌های جدول اول

انجام گیرد، اصطلاحاتی که بسامد بالاتری دارند شانس هم‌رخدادی بیشتری با سایر اصطلاحات پیدا خواهند کرد (۹). به این جهت اعداد به مقیاسی نسبی در بسته‌ی بین صفر و یک تبدیل شدند.

سمت چپ مربوط به دوره زمانی ۹۹-۱۹۹۰ و سمت راست مربوط به دوره زمانی ۲۰۰۸-۲۰۰۰ است. خطوطی که در نقطه‌ی بین فاصله‌ی ۱۰ و ۱۵ بر نمودارها عمود شده است، خط شاخص تفسیر می‌باشد که توسط متخصص موضوعی رسم شده است. مستطیل‌ها نشان می‌دهند یازده دسته به وجود آمده بر اساس این خوشه‌بندی، در هر دوره‌ی زمانی شامل چه توصیفگرهایی هستند.

این کار با این آگاهی صورت گرفت که روش Ward مبتنی بر فاصله‌ی اقلیدسی است (۶) و به کارگیری آن با سایر روش‌های اندازه‌گیری مانند روش کوسینوسی به طور دقیق نتیجه‌اش آن خوشه‌بندی نیست که در روش Ward همراه با سنجش فاصله‌ی اقلیدسی صورت می‌گیرد. خوشه‌بندی Ward با روش سنجش کوسینوسی منجر به تشخیص توصیفگری نخواهد شد که بیشترین فاصله را با سایر توصیفگرها دارد. چیزی که در اصطلاح رانت (Runt) نامیده می‌شود.

برای سنجش میزان ارتباط بین اصطلاحات، می‌توان بسامد هم‌رخدادی آن‌ها را هم به صورت مطلق و هم به صورت نسبی در نظر گرفت. اما اگر این سنجش را بر پایه‌ی بسامد مطلق



شکل ۲: خوشه‌بندی سلسله‌مراتبی ۳۸ توصیفگر مربوط به روان‌شناسی ازدواج در ادبیات زیست‌پزشکی

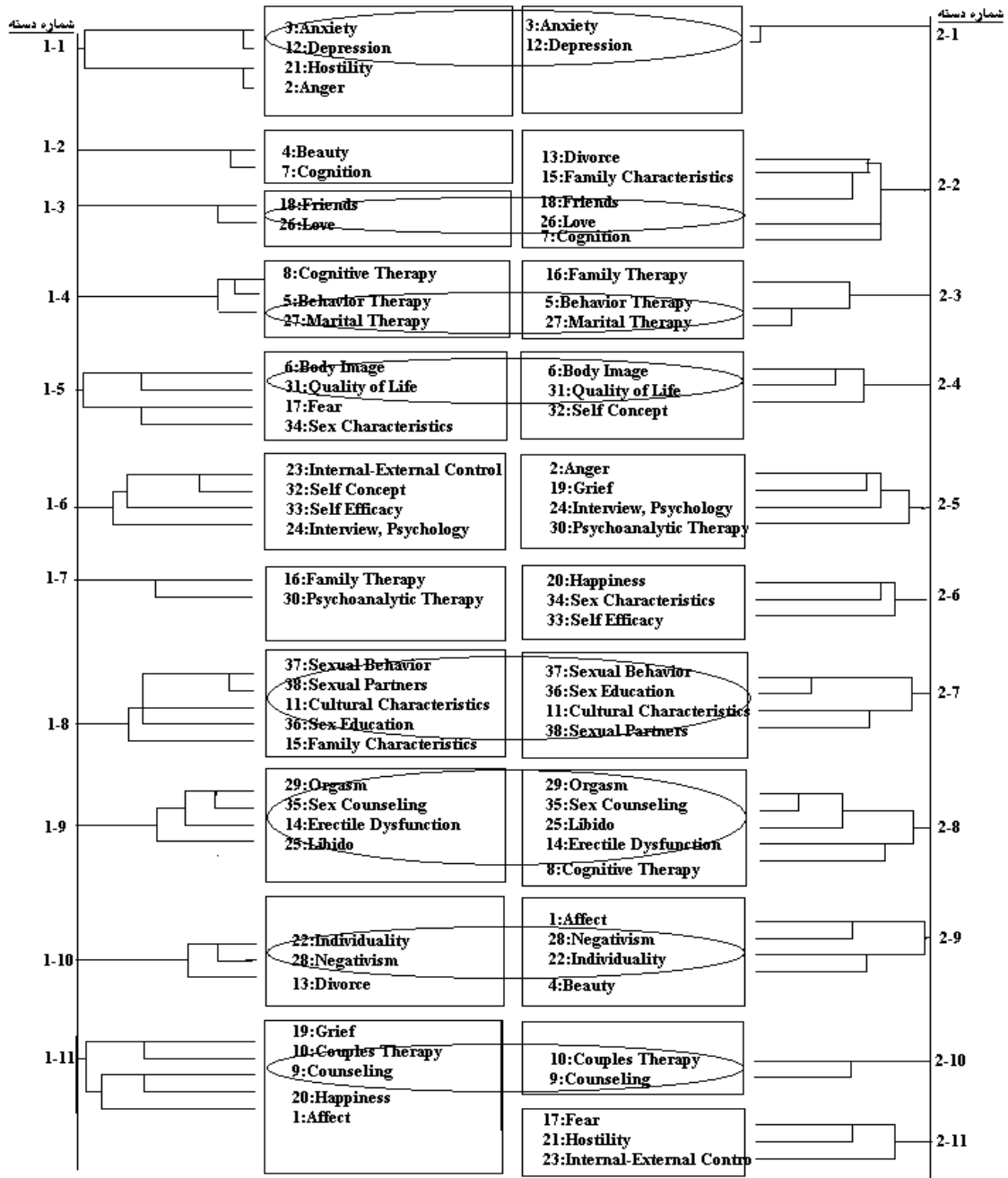
یافته‌ها

ماتریس ۳۸ × ۳۸ انجام شده است. ماتریس اول و دوم به ترتیب متعلق به هم‌رخدادی توصیفگرهای مربوط به دوره‌های زمانی ۹۹-۱۹۹۰ و ۲۰۰۸-۲۰۰۰ بود (شکل ۲ و ۳).

خوشه‌بندی سلسله‌مراتبی توصیفگرها، به روش Ward و بر اساس محاسبه‌ی فاصله‌ی (شباهت) کوسینوسی آن‌ها برای دو

دسته بندی توصیفگرهای مربوط به دوره زمانی ۱۹۹۰-۱۹۹۹

دسته بندی توصیفگرهای مربوط به دوره زمانی ۲۰۰۰-۲۰۰۸



شکل ۳: مقایسه‌ی یازده گروه از توصیفگرهای حاصله بر اساس خوشه‌بندی سلسله‌مراتبی

پژوهشگر صورت می‌گیرد. هدف پژوهش حاضر، بررسی سیر تحول رابطه‌ی بین موضوعات از لحاظ هم‌رخدادی در متون مربوط به روان‌شناسی ازدواج است که در حوزه‌ی زیست پزشکی (به طور خاص در Medline) مورد توجه قرار گرفته است. به همین جهت دسته‌های به وجود آمده، بر خلاف واژه‌های کنترل شده، مانند تزاروس، نشانگر روابط عام و خاص بین اعضای آن‌ها نبوده است. این دسته‌ها روابط پنهانی را نشان می‌دهند که بین موضوعات وجود دارد و از طریق روش‌های ذهن‌محور قابل کشف نیست.

برای تحلیل روابط پنهان بین موضوعات، ابتدا سی و هشت توصیفگر مورد تحقیق با توجه به ارتباط مفهومی بین آن‌ها به پنج حوزه‌ی موضوعی تقسیم شده است (جدول ۳):

در ادامه، تحلیل روابط موضوعی روان‌شناسی ازدواج در ادبیات زیست پزشکی، بر اساس نمودار شکل ۳ و در چار چوب پنج حوزه‌ی نامبرده در جدول ۳ صورت گرفت. این عمل علاوه بر مشخص کردن این روابط، نشان می‌دهد که به عنوان مثال یک متخصص امر روان‌شناسی یا اطلاع‌رسانی چگونه می‌تواند از نتایج حاصل از خوشه‌بندی بهره برد. باید توجه داشت که منظور از شماره‌ی دسته، شماره‌ی نسبت داده شده به هر یک از خوشه‌ها در شکل ۳ است:

حوزه‌ی مشاوره و درمان:

در دوره‌ی اول (۹۹-۱۹۹۰) پنج موضوع درمانی وجود داشت که با هیچ موضوع دیگر در یک دسته قرار نگرفته بود یا به عبارتی، با موضوعات غیر درمانی هم‌جوار نبوده‌اند. این پنج روش درمانی عبارت است از شناخت درمانی، رفتار درمانی، ازدواج درمانی (در دسته‌ی ۲-۱)، خانواده درمانی و روان درمانی (در دسته ۶-۱). می‌توان چنین نتیجه گرفت که در این پنج موضوع، بیشتر ماهیت یا ارتباط آن روش‌ها با یکدیگر مورد تأکید بود. در دوره‌ی بعدی (۲۰۰۸-۲۰۰۰) هنوز سه مورد از این پنج موضوع با موضوعات غیر درمانی هم‌جوار نبوده‌اند (خانواده درمانی، رفتار درمانی و ازدواج درمانی در دسته‌ی ۳-۲) از طرفی مشاوره و زوج درمانی که در دهه‌ی ۱۹۹۰ با

روان‌شناس متخصص پس از مشاهده‌ی نتایج حاصل از این دو تحلیل، محل عمود شدن خط شاخص تفسیر را بر هر یک از این دو نمودار درختی تعیین کردند. این خطوط شاخص نشان داد که بهترین حالت در هر یک از این دو نمودار تقسیم توصیفگرها به یازده دسته است. بر این اساس، از هر نمودار یازده دسته‌ی توصیفگر استخراج شد. برای مقایسه‌ی دسته‌های حاصل شده در دو دوره‌ی زمانی، دسته‌هایی که توصیفگرهای مشترک بیشتری داشتند، در کنار یکدیگر جفت شدند. از طریق این روش، مشترکات و تفاوت‌های آن‌ها به راحتی به تصویر کشیده شد.

سمت چپ و راست به ترتیب مربوط به دوره‌های زمانی ۱۹۹۰-۱۹۹۰ و ۲۰۰۸-۲۰۰۰ است. مستطیل‌ها نشان دهنده این است که چه توصیفگرهایی با هم در یک دسته قرار گرفته‌اند. بیضی‌های خاکستری رنگ مشخص می‌سازند چه توصیفگرهایی در هر دو دوره‌ی زمانی همواره در کنار هم، در یک دسته قرار داشته‌اند. نمودارهای جنب مستطیل‌ها نیز برش‌هایی از نمودارهای تصویر است که ارتباط سلسله مراتبی هر توصیفگر را با توصیفگر دیگر در هر دسته به تصویر می‌کشد. شماره‌هایی که در کنار هر توصیفگر آمده است، شماره‌ی ردیف آن‌ها در حالت تنظیم بر اساس الفبایی است. شماره‌های موسوم به شماره‌ی دسته، شماره‌ی ترتیب هر دسته در هر دوره است. این شماره از طرف پژوهشگران داده شده است تا در بحث و تحلیل نتایج، ارجاع به دسته‌ها امکان‌پذیر باشد.

مقایسه‌ی دسته‌های به نسبت متناظر هر دوره، نشان می‌دهد که در مجموع ۲۰ توصیفگر از ۳۸ مورد، در طی دو دوره‌ی مورد پژوهش در هشت دسته جداگانه همیشه هم‌جوار بوده‌اند (دسته‌های موجود در بیضی‌ها). اما فاصله‌ی هم‌جواری آن‌ها در هر یک از دو دوره در اغلب دسته‌ها متفاوت بود. به طور مثال دو توصیفگر Friends و Love یکی از این دسته‌ها را تشکیل می‌دهند که در دوره‌ی زمانی اول ضریب فاصله‌ی آن‌ها کمتر بود و در دوره‌ی دوم بسیار افزایش یافته است.

بحث

تحلیل نتایج حاصل از خوشه‌بندی بنابر هدف پژوهش و

موضوعات احساسی و عاطفی در یک دسته قرار گرفته بودند (دسته‌ی ۱۰-۱)، در دوره‌ی بعدی، بدون هم‌جواری شدن با موضوعی دیگر، باز به تنهایی در یک دسته (دسته‌ی ۹-۲) در کنار هم قرار گرفته‌اند. یعنی این دو روش که مکمل یکدیگر هستند، دیگر در دوره‌ی بعدی به طور برجسته برای درمان موارد خاصی مورد توجه قرار نگرفته‌اند.

جدول ۳: تقسیم‌بندی و هشت توصیفگر مورد بررسی به پنج حوزه‌ی موضوعی

توصیفگرها	Descriptors	حوزه موضوعی
شناخت درمانی	Cognitive Therapy	مشاوره و درمان
رفتار درمانی	Behavior Therapy	
ازدواج درمانی	Martial Therapy	
خانواده درمانی	Family Therapy	
روان درمانی	Psychoanalytic Therapy	
زوج درمانی	Couples Therapy	
مصاحبه، روان‌شناسی مشاوره	Interview, Psychology Counseling	
مشاوره جنسی	Sex Counseling	
رفتار جنسی	Sexual Behavior	جنسی
شرکای جنسی	Sexual Partners	
آموزش جنسی	Sex Education	
اورگاسم	Orgasm	
اختلال در نعوذ	Erectile Dysfunction	
شور جنسی	Libido	
ویژگی‌های جنسیتی	Sex Characteristics	خصوصیات فردی و اجتماعی
ویژگی‌های فرهنگی	Cultural Characteristics	
ویژگی‌های خانوادگی	Family Characteristics	
کیفیت زندگی	Quality of Life	
فردیت	Individuality	
خود پنداره	Self Concept	
خود بسندگی	Self Efficacy	
کنترل درونی و بیرونی	Internal-External Control	
اضطراب	Anxiety	احساسی - عاطفی - هیجانی - خلقی - ادراکی
افسردگی	Depression	
خصومت	Hostility	
خشم	Anger	
زیبایی	Beauty	
ادراک	Cognition	
منفی‌گرایی	Negativism	
اندوه	Grief	
شادکامی	Happiness	
عاطفه	Affect	
تصویر بدن	Body Image	
ترس	Fear	
دوستان	Friends	دوستی - عشق - جدایی
عشق	Love	
طلاق	Divorce	

با موضوعات دیگر حفظ کرده است: ۱- خصوصیات فرهنگی که در دوره‌های اول و دوم با دسته‌ی موضوعات مربوط به رفتارهای جنسی در ارتباط بود (دسته‌ی ۱-۸ و ۲-۷). ۲- فردیت که همواره با منفی‌گرایی (دسته‌ی ۱-۱۰ و ۲-۹) در حول یک محور قرار می‌گیرد. این دو موضوع اخیر، در دوره‌ی اول با جدایی و در دوره‌ی دوم همراه با یک خصوصیت فردی دیگر (زیبایی)، در کنار موضوع عاطفه، یک بخش از ادبیات مربوطه را تشکیل داده‌اند.

خصوصیات فردی و اجتماعی از موضوعاتی هستند که اغلب کمترین ثبات را از جنبه‌ی هم‌جواری با موضوعات سایر حوزه‌ها و حتی درون حوزه‌ی خودشان با یکدیگر داشته‌اند.

حوزه‌ی احساسی-عاطفی-هیجانی-خلقی-ادراکی:

اکثر موضوعات این حوزه در دوره‌ی زمانی اول در دو دسته (دسته‌ی ۱-۱ شامل اضطراب، افسردگی، خصومت و خشم و دسته‌ی ۱-۱۰ شامل اندوه، شادکامی و عاطفه) جای گرفته‌اند. آن چه مشاهده می‌شود، بر خلاف تصور، چهار موضوع اول در رابطه با هیچ یک از موضوعات درمانی و مشاوره‌ای نیستند. در جستجوی جداگانه‌ای که برای بررسی دلیل این موضوع در Medline انجام شد، مشخص گردید که این موضوعات، به ویژه موضوع اضطراب و افسردگی، بیشتر با درمان دارویی و مواردی از این قبیل در ارتباط می‌باشد که این موضوع در انتخاب موضوعات مد نظر روان‌شناس متخصص قرار نگرفته است. سلسله مراتب ارتباطات در دسته‌ی دیگر (دسته‌ی ۱-۱۰) نشان می‌دهد که موضوع اندوه با زوج درمانی و شادکامی و همچنین عاطفه با مشاوره، مجاورت بیشتری داشته‌اند. به هر حال موضوعات این حوزه در دوره‌ی اول مورد مطالعه، پیوند مستقیمی با درمان، مصاحبه و مشاوره داشته است که در هر مورد روشی خاص استفاده شده است.

در دوره‌ی زمانی بعد (۲۰۰۸-۲۰۰۰) بیشتر موضوعات این حوزه در سه دسته قرار می‌گیرد. یک دلیل بر افزایش دسته‌ی دیگر، نسبت به دوره‌ی قبل، ارتباط شدید بین اضطراب و افسردگی است که در هر دو دوره فاصله‌ی بین این دو موضوع

هم‌جواری روش‌های درمانی با یکدیگر و هم‌جواری نشدن اکثر این موضوعات با سایر موضوعات، مؤید این نکته است که غالب روش‌های درمانی با وجود قدمت همواره از حیث اصالت و کارکردشان در زمینه‌ی مسایل ازدواج، مورد بحث روان‌شناسان هستند و هنوز روش‌های درمانی خاصی به طور قطع نتوانسته برای درمان مسایل روانی مربوط به ازدواج پذیرفته شوند. چرا که از دیدگاه روان‌شناختی، به تعداد مشکلات افراد، شیوه‌های درمانی خاصی نیاز است. از این میان، تنها موضوع مشاوره جنسی است که بنابر طبیعتش همواره در کنار موضوعات مربوط به مقاربت و همبستری، به ویژه اورگاسم، قرار گرفته است (دسته‌ی ۱-۹ و ۲-۸).

حوزه‌ی جنسی:

در طول زمان، با ثبات‌ترین حوزه در مباحث مربوط به روان‌شناسی ازدواج در ادبیات پزشکی، حوزه‌ی مسایل جنسی است که در هر دو دوره‌ی زمانی مورد مطالعه، موضوعاتشان همواره هم‌جواری با یکدیگر را حفظ کرده‌اند (دسته‌ی ۱-۸، ۱-۹، ۲-۷ و ۲-۸). علاوه بر آن، موضوعات جنسی همیشه در دو دسته مورد توجه بوده‌اند: دسته‌ی موضوعات رفتار جنسی (دسته‌ی ۱-۸ و ۲-۷) و دسته‌ی موضوعات مقاربتی (دسته‌ی ۱-۹ و ۲-۸). در مورد اول، رفتار جنسی در شرکای جنسی از جنبه‌ی خصوصیات اجتماعی، در کنار آموزش جنسی، یک محور از ادبیات مورد پژوهش را تشکیل داده است.

در مورد دوم (دسته‌ی موضوعات مقاربتی)، تنها نکته‌ی مورد توجه هم‌جواری شدن موضوع شناخت درمانی با مسایل مقاربتی در دوره‌ی زمانی ۲۰۰۸-۲۰۰۰ است. در این دوره همانند دوره‌ی قبل، در سلسله مراتب ارتباطی، مشاوره جنسی با اورگاسم ارتباط بیشتری داشته است و شناخت درمانی به عنوان یک روش درمانی در مجاورت موضوع مشکل نمودن قرار گرفته است.

حوزه‌ی خصوصیات فردی و اجتماعی:

در ادبیات مورد بررسی از هشت موضوع مربوط به حوزه‌ی خصوصیات فردی و اجتماعی، دو مورد پیوستگی خود را همواره

خانوادگی بود. سرانجام، در فاصله‌ای دورتر عشق و ادراک، به عنوان دو موضوع نزدیک به هم، با جدایی در ارتباط قرار می‌گیرد.

نتیجه‌گیری

در این پژوهش نشان داده شد که چگونه می‌توان دو نمودار سلسله مراتبی را که در یک زمینه‌ی موضوعی خاص از دو مجموعه‌ی داده‌ی جداگانه استخراج شده است، بدون تکیه بر الگوهای ریاضی و آماری خاص مقایسه و از آن نتیجه‌گیری کرد. این روش تحلیلی روشن ساخت که در حیطه‌ی اطلاع‌رسانی می‌توان از پایگاه داده‌های کتاب‌شناختی، مانند Medline، علاوه بر جستجو و بازیابی مدارک متناسب با نیاز اطلاعاتی استفاده‌های دیگری نیز کرد. با تحلیل موضوعات متناسب به آن مدارک، به ویژه توصیف‌گرها، می‌توان به بعضی از سؤالات مرجع پاسخ گفت. این کار از طریق روش‌های سنجش شباهت و فاصله‌ی بین توصیف‌گرهای مدارک بازیابی شده است و خوشه‌بندی آن‌ها امکان‌پذیر می‌باشد. به طور مثال اگر شخصی بخواهد بداند که چه روش‌های درمانی برای مسایل عاطفی، همراه با چه ویژگی‌های اجتماعی و فردی، در ادبیات مربوط به ازدواج مورد بحث قرار گرفته است، این روش می‌تواند به این سؤال پاسخ بگوید. این در حالی است که یافتن پاسخ چنین سؤالاتی به راحتی از منابع مرجع امکان‌پذیر نیست. این از توانایی‌های روش خوشه‌بندی سلسله مراتبی است که در ادبیات مربوط به کار مرجع به آن توجه نشده است.

کمترین میزان را نسبت به تمامی موضوعات داشته است و این رابطه‌ی شدید به گونه‌ای بوده است که در دوره‌ی بعدی یک بخش جداگانه در ادبیات مربوط به روان‌شناسی ازدواج در زیست پزشکی به این دو موضوع تعلق گرفته است. دو دسته‌ی دیگر به ترتیب به این گونه است: دسته‌ی ۵-۲ خشم و اندوه، همراه با روان‌درمانی و مصاحبه، دسته‌ی ۱۰-۲ شامل موضوعات ترس، خصومت بدون ارتباط با موضوعات مشاوره‌ای و درمانی است. موضوعات این حوزه تمایل زیادی به مجاور بودن با یکدیگر دارند. به گونه‌ای که در معمول هر یک از موضوعات این حوزه، همراه با موضوع یا موضوعات دیگری از همین حوزه هم‌جوار است.

حوزه‌ی دوستان، عشق و جدایی:

در این حوزه از سه موضوع مورد بررسی (دوستان، عشق و جدایی)، دو مورد اول همواره در دو دوره‌ی زمانی مورد مطالعه با هم هستند که در دوره‌ی اول فاصله‌ی مجاورتی بین این دو بسیار کم است و در دوره‌ی بعد افزایش می‌یابد. نکته‌ی دیگر این که در دوره‌ی زمانی دوم، هر سه موضوع در یک دسته قرار می‌گیرند.

اگر در این حوزه، موضوع جدایی (طلاق) به عنوان یک موضوع محوری در نظر گرفته شود، این نتیجه حاصل می‌شود که جدایی در دهه‌ی آخر قرن بیستم از دریچه‌ی دو موضوع پیوسته، همراه فردیت و منفی‌گرایی در ادبیات زیست پزشکی وارد شده است. در دهه‌ی اول قرن بیست و یکم توجه به مسأله طلاق همراه با موضوعات دوستان و خصوصیات

References

1. Jacob T. Family interaction and psychopathology: theories, methods, and findings. New York: Springer; 1987.
2. Hamilton GV. A research on marriage's. New York: Lear; 1948.
3. Fincham FD, Bradbury TN. Psychology and the study of marriage. New York: Guilford Press; 1990.
4. Mack R, Hehenberger M. Text-based knowledge discovery: search and mining of life-sciences documents. Drug Discov Today 2002; 7(11 Suppl): S89-S98.
5. Borgen FH, Barnett CB. Applying cluster analysis in counseling psychology research. Journal of Counseling Psychology 1987; 34(4): 456-68.
6. Korenius T, Laurikkala J, Juhola M. On Principal component analysis, cosine and euclidean measures in information retrieval. Information Sciences 2007; 177(22): 4893-905.

7. Miyamoto S. Proximity Measures for term based on fuzzy neighborhoods in document sets. *International Journal of Approximate Reasoning* 2003; 34(2-3): 181-99.
8. Buscaldi D, Rosso P, Alexandrov M, Ciscar AJ. Sense cluster based categorization and clustering of abstracts. In: Buscaldi D, Editor. *Computational linguistics and intelligent text proceeding*. Berlin: Springer, 2006. p. 547-50.
9. Chung YM, Lee JY. A corpus-based approach to comparative evaluation of statistical term association measures. *Journal of the American Society for Information Science and Technology* 2001; 52(4): 283-96.
10. Yasukawa M, Yokoo H. Related terms clustering for enhancing the comprehensibility of web search results. In: Wagner RR, Revell N, Pernul G, Editors. *Database and expert systems applications: 18th international conference*. Berlin: Springer, 2007. p. 359-68.
11. Tao L, Zhu S, Ogihara M. Hierarchical document classification using automatically generated hierarchy. *Journal of Intelligent Information Systems* 2007; 29(2): 211-30.
12. Koller D, Sahami M. Hierarchically classifying documents using very few words. *Proceedings of the 14th International Conference on Machine Learning*; 1997 Jul; Nashville, Tennessee. San Francisco, CA: Morgan Kaufmann Publishers Inc; 1997. p. 170-8.
13. Swanson DR, Smallheiser NR. Implicit text linkages between medline records: Using arrowsmith as an aid to scientific discovery. *ERIC* 1999; 48(1): 48-59.
14. Janssens F, Glaenzel W, De Moor B. A hybrid mapping of information science. *Scientometrics* 2008; 75(3): 607-31.
15. Stephen T. Concept analysis of the communication literature on marriage and family 110. *Journal of Family Communication* 2001; 1(2): 91-110.
16. Jain AK, Murty MN, Flynn PJ. Data clustering: a review and computing surveys. *ACM* 1999; 31(3): 295-323.
17. Ward JH. Hierarchical grouping to optimize an objective function. *Journal of American Statistical Association* 1963; 58(301): 236-44.
18. Nenadic G, Spasic I, Ananiadou S. Term clustering using a corpus-based similarity measure. In: Sojka P, Kopecek I, Pala K, Editors. *Text, speech, and dialogue: 5th International Conference, TSD 2002*. Berlin: Springer, 2002. p. 151-4.

Document-Term Clustering: Proximity of Subjects Correspond with Psychology of Marriage in Biomedicine Literature during the Years “1990-99” and “2000-2008”*

Mohammad Tavakolizadeh Ravari, PhD¹; Maryam Nejabatian²

Abstract

Introduction: The study on psychology of marriage rooted in the social issues. The field has also found its way to biomedicine. The current research, based on the document-term clustering methods, focus on term-proximity of documents that were related to the field of psychology of marriage in biomedicine literature.

Methods: The co-occurrence frequency of terms within the two different corpuses of documents resulted in two proximity matrixes. The corpuses contained sequentially the documents published during the years “1990-99” and “2000-2008”. Those documents were returned by two separate searches in Medline. The cosine method was established to determine the proximity of terms. Besides, the terms were clustered by the Ward’s method.

Results: The measurement of term-proximity resulted in two graphs of hierarchical clusters. The comparison revealed that twenty terms up to thirty-eight kept permanently their proximity with one or more other terms during the two periods of study adult and 21(60%) prenatal had a correct sequence.

Conclusion: Dividing correspond subjects with psychology of marriage into five areas prepares the opportunity to compare the two hierarchical graphs and find the relationship between and within those areas. Consequently, document-term clustering assists to answer some reference queries.

Keywords: Psychology; Marriage; Classifications; Medicine in Literature; Medline; Subject Headings.

Type of article: Original Article

Received: 11 Dec, 2008

Accepted: 27 Aug, 2009

Citation: Tavakolizadeh Ravari M, Nejabatian M. **Document-Term Clustering: Proximity of Subjects Correspond with Psychology of Marriage in Biomedicine Literature during the Years “1990-1999” and “2000-2008”**. *Health Information Management* 2010; 7(2): 186.

*This article resulted from independent research.

1. Assistant Professor, Library and Information Science, Yazd University, Yazd, Iran. (Corresponding Author)

Email: tavakoli@vazduni.ac.ir

2. Msc, Psychology, National Youth Organization, Tehran, Iran.