

چارچوبی برای استفاده از دانش مشتریان در طراحی کاتالوگ‌های الکترونیکی

نگار قنبری¹ و محمدرضا غلامیان^{2*}

¹ دانش آموخته کارشناسی ارشد مهندسی فناوری اطلاعات (تجارت الکترونیکی) - دانشکده مهندسی صنایع دانشگاه علم و صنعت

² استادیار گروه مهندسی فناوری اطلاعات (تجارت الکترونیکی) - دانشکده مهندسی صنایع دانشگاه علم و صنعت

(تاریخ دریافت 88/8/18، تاریخ دریافت روایت اصلاح شده 89/8/3، تاریخ تصویب 90/1/17)

چکیده

تحقیقات نشان داده است که در میان موضوعات مختلف مطرح در تجارت الکترونیکی، مدیریت اثربخش مشتریان در کنار استفاده از دانش آنها از مواردی است که می‌تواند به مزیت رقابتی منجر شود. کاوش دانش مشتریان برای کمک به مدیران برای در پیش گرفتن خط مشی‌های تبلیغاتی بهتر، با بهره‌گیری از سیستمی که بتواند الگوهای رفتار مشتریان را از پایگاه داده‌ها استخراج کند، امکان‌پذیر است. خروجی چنین سیستمی را می‌توان در قالب خط مشی‌های تبلیغاتی در کار طراحی کاتالوگ‌های الکترونیکی به کار گرفت. در این مقاله یک چارچوب برای طراحی کاتالوگ‌های الکترونیکی بر اساس دانش مشتریان ارائه می‌شود. این چارچوب از الگوریتم جدیدی برای کاوش چندسطحی الگوهای متوالی به نام DVIW استفاده می‌کند و در مورد یک شرکت فروشنده تجهیزات رایانه‌ای به کار گرفته شده است. بررسی‌ها نشان می‌دهد که دانش نهایی به دست آمده توسط آن در مقایسه با سایر چارچوب‌های معرفی شده در این زمینه کامل‌تر است.

واژه‌های کلیدی: کاتالوگ‌های الکترونیکی، الگوهای متوالی، دانش مشتریان

مقدمه

مشتری بیشترین کاربرد را دارند [2، 3]. از آنجایی که در الگوهای متوالی زمان نقش تعیین‌کننده‌ای دارند، می‌توانند در پیش‌بینی خریدهای بعدی مشتری به طور گسترده به کار گرفته شود [1].

در این تحقیق، یک چارچوب برای طراحی کاتالوگ‌های الکترونیکی بر اساس دانش مشتریان معرفی می‌شود. الگوهای متوالی ابزاری است که این چارچوب برای استخراج دانش مشتریان از آن بهره می‌گیرد که در ادامه شرح داده می‌شود.

ساختار کلی مقاله به این ترتیب است: در بخش بعد مروری بر ادبیات مدل‌ها و چارچوب‌ها و روش‌های داده‌کاوی مورد استفاده در استخراج دانش مشتری انجام می‌گیرد. بخش سوم، چارچوب جدید استفاده از دانش مشتری در طراحی کاتالوگ‌های الکترونیکی را معرفی می‌کند. بخش چهارم مقاله به معرفی شرکت مورد مطالعه و نتایج تجربی پیاده‌سازی چارچوب پیشنهادی در آن شرکت می‌پردازد. بخش پنجم به بررسی دانش به دست

فناوری اطلاعات ابزارهای بسیاری را برای مشتریان و شرکت‌ها برای کمک به مسائل تصمیم‌گیری فراهم کرده است. برای مثال، مشتریان با استفاده از اینترنت به راحتی می‌توانند به اطلاعات محصولات بیشماری از فروشندگان مختلف دسترسی داشته باشند که این اطلاعات بر تصمیم‌گیری خرید آنها تأثیر خواهد گذاشت. علاوه بر این شرکت‌ها نیز می‌توانند با جمع‌آوری و تحلیل اطلاعات مشتریان، تصمیمات بهتری را در زمینه بازاریابی در پیش بگیرند.

در گذشته، محققان برای بررسی رفتار مشتریان از تحلیل‌های آماری استفاده می‌کردند که امروزه جای خود را به روش‌های داده‌کاوی داده است. داده‌کاوی از فناوری‌های پرکاربرد برای کشف دانش بالقوه مشتریان از پایگاه داده‌های کسب و کار برای کمک به تصمیم‌گیری است [1]. از بین روش‌های مختلف داده‌کاوی شامل قوانین تلازمی، خوشه‌بندی، طبقه‌بندی و ... قوانین تلازمی و الگوهای متوالی در تحلیل داده‌های خرید

گرفتن تصمیمات کسب و کار نیز هدفی است که این مقاله آن را دنبال می‌کند.

علاوه بر تعاریف متعدد از مدیریت دانش مشتری، مدل‌ها و چارچوب‌هایی نیز در این زمینه ارائه شده است. سالومن و همکاران [4] چارچوبی برای پیاده‌سازی CRM مبتنی بر دانش ارائه دادند که از چهار بخش استراتژی‌ها، فرآیندها، سیستم‌ها و مدیریت تغییر تشکیل شده است و در بخش سیستم‌های این چارچوب، هدف ایجاد یک مخزن دانش یکپارچه در محدوده سازمان دنبال می‌شود. سو و همکاران [7] مدلی مفهومی برای $e\text{-CKM}^1$ ارائه دادند که از دانش مشتری برای نوآوری در توسعه محصول استفاده می‌کرد. در نظر گرفتن بخش‌بندی بازار و جداسازی الگوهای مشتریان با توجه به آن از نکاتی است که در این مدل وجود دارد. لیو و چن [8] به طور خاص به استفاده از دانش مشتری در طراحی کاتالوگ‌های اینترنتی پرداخته‌اند و وظایف را شامل تشکیل پایگاه داده، وظایف داده‌کاوی، تحلیل بخش‌بندی، جذب دانش، بازاریابی مبتنی بر کاتالوگ و تبلیغات فروش برشمرده‌اند. در این مقاله، گرچه داده‌کاوی به عنوان مرحله‌ای اصلی برای استخراج دانش مشتری در نظر گرفته شده و الگوها در سه سطح خود محصول، برند محصول و دپارتمان محصول به دست آمده است، اما اشاره دقیقی به الگوریتم مورد استفاده وجود ندارد. لین و هنگ [9] نیز از روش مشابه مورد ذکر شده برای ایجاد کاتالوگ اینترنتی استفاده کردند. خروجی هر دو مقاله، الگوهایی برای طراحی کاتالوگ‌های الکترونیکی بود. چنگچن و همکاران [10] یک سیستم پشتیبانی تصمیم‌گیری برای تبلیغات فروش شخصی‌سازی شده، ایجاد کردند که به دست آوردن الگوها در سه سطح تمام مشتریان/خوشه‌های مشتریان و هر مشتری از ویژگی‌های بارز آن است.

دانش مشتری و روش‌های داده‌کاوی

روش‌های داده‌کاوی گوناگون خوشه‌بندی، دسته‌بندی، قوانین تلازمی، درخت‌های تصمیم‌گیری و ... را می‌توان به دو دسته کلی توصیفی و پیش‌بینی‌کننده تقسیم کرد. لیو و چن [8] و لین و هنگ [9] در چارچوب‌های معرفی‌شده برای طراحی کاتالوگ‌های اینترنتی به استفاده از قوانین تلازمی و الگوهای متوالی اشاره کرده‌اند. چنگچن و همکاران [10] نیز در سیستم پشتیبانی تصمیم‌گیری خود

آمده از پیاده‌سازی چارچوب اختصاص دارد و در نهایت جمع‌بندی و پیشنهادات برای تحقیقات بعدی ارائه می‌شود.

مروری بر ادبیات

همان‌طور که اشاره شد، هدف مقاله معرفی چارچوب برای استفاده از دانش مشتری در طراحی کاتالوگ‌ها است، اما استفاده از دانش مشتری، خود مستلزم استخراج دانش از داده‌ها و بررسی روش‌های داده‌کاوی مناسب برای این کار است. بنابراین ادبیات موضوع را در دو محور مدیریت دانش مشتری و روش‌های داده‌کاوی برای استخراج دانش مشتری دنبال می‌کنیم.

مدیریت دانش مشتری

برای مدیریت دانش مشتری تعاریف زیادی در مقالات ارائه شده است. گیبرت و همکاران [2] مدیریت دانش مشتری را یک فرآیند استراتژیک معرفی کرده‌اند که در آن شرکت‌ها مشتریان را از دریافت‌کننده منفعل کالا و خدمات به قدرتمندی شرکای دانش می‌رسانند و 5 دسته یادگیری مشارکتی تیمی، نوآوری دوجانبه، انجمن‌های خلاقیت، مدیریت مشترک سرمایه معنوی، تولیدکننده-مصرف‌کننده را برای دانش مشتری در نظر می‌گیرند. گیبرت و همکاران [3] هدف از مدیریت دانش مشتری را با مباحث مدیریت دانش مشتری و مدیریت دانش یکسان دانسته و این هدف را به صورت ایجاد بهبود دائم برای مشتریان تعریف کردند. سالومن و همکاران [4] ضمن تعریف سه دسته برای دانش مشتریان (دانش از/برای/درباره مشتریان)، هدف آن را گسترش توانایی‌های مرتبط با مشتری در سازمان‌ها دانسته‌اند. پاکوت [5] نیز مدیریت دانش مشتری را به صورت فرآیندی که سازمان برای مدیریت شناسایی، جذب و به کارگیری داخلی دانش مشتری به کار می‌گیرد و در خلال آن سازمان و مشتریان آن با یکدیگر برای ترکیب دانش فعلی برای ایجاد دانش جدید کار می‌کنند، تعریف کرد. لویز و مولینا [6] مدیریت دانش مشتریان را کاربرد ابزارها و تکنیک‌های مدیریت دانش برای پشتیبانی از تبادل دانش بین سازمان و مشتریان دانسته که به شرکت امکان گرفتن تصمیمات کسب و کار را می‌دهد. استفاده از دانش مشتری برای

کاتالوگ‌های الکترونیکی، الگوریتم مورد استفاده برای کاوش چندسطحی الگوهای متوالی نیز تشریح می‌شود.

معرفی چارچوب

با توجه به جمع‌بندی مدل‌ها و چارچوب‌ها در مقالات که در بخش قبل مطرح شد، چارچوب نهایی برای طراحی کاتالوگ‌های الکترونیکی بر اساس دانش مشتریان پیشنهاد می‌شود. (شکل 1)

در این چارچوب، پایگاه‌های داده برای دسته‌بندی چندسطحی محصولات و خوشه‌بندی مشتریان دیده می‌شود. دسته‌بندی محصولات، بر اساس پایگاه داده محصولات و نظر خبرگان در چندین سطح انجام می‌گیرد. برای مثال اگر با توجه به پایگاه داده محصولات در سطح اول، 5 دسته برای محصولات موجود است، در سطح دوم، بنا به نظرات خبرگان دسته‌های 1 و 3 ممکن است در درون دسته‌ای جدید به نام A قرار بگیرند و تا آخرین سطح در نظر گرفته شده برای محصولات، این خاصیت ادامه پیدا می‌کند. در نظر گرفتن این دسته‌بندی پویا در داده‌کاوی می‌تواند به کشف الگوهای جالبی در رفتار خرید مشتریان منجر شود.

خوشه‌بندی مشتریان نیز با استفاده از ویژگی‌هایی که خبره حوزه تعیین می‌کند، انجام می‌شود. در نهایت با توجه به دسته‌بندی چندسطحی محصولات و خوشه‌بندی انجام شده، الگوهای خرید در سطوح متفاوت دسته‌بندی محصول برای همه مشتریان، خوشه‌های مشتریان و هر مشتری به دست می‌آید. سپس با دخیل کردن نظرات خبره حوزه در زمینه انتخاب الگوها و تعیین استراتژی‌ها، الگوها برای قرار گرفتن در پایگاه دانش طراحی کاتالوگ‌های الکترونیکی نهایی می‌شوند.

معرفی الگوریتم جدید DVIw6

استفاده از چارچوب معرفی شده بدون داشتن ابزاری که فرآیند استخراج دانش مشتریان را از پایگاه‌های داده امکان‌پذیر کند، میسر نیست. بنابراین در ادامه به معرفی الگوریتم DVIw که در این چارچوب مورد استفاده قرار می‌گیرد، می‌پردازیم. این الگوریتم جدید برای کاوش چندسطحی الگوهای متوالی همانند الگوریتم GSP بر اساس روش تولید و تست کاندیدها عمل می‌کند، اما کاوش چندسطحی الگوها را به نحو کارآتری نسبت به

از این روش‌ها بهره گرفته‌اند. پرینز و پل [11] نیز برای پیش‌بینی رفتار مشتری از الگوهای متوالی استفاده کرده‌اند. در همه این مقالات، هدف، استفاده از الگوهای متوالی استخراج دانش مشتری بوده است. در برخی مقالات که به بخش‌بندی بازار قبل از به دست آوردن الگوها اشاره شده است، روش خوشه‌بندی روش غالب برای این کار بوده است [7, 10].

از آنجایی که در مقالات، اشاره دقیقی به الگوریتم‌های مورد استفاده برای کاوش الگوها نشده است، در ادامه به الگوریتم‌های موجود برای کاوش الگوهای متوالی و به طور خاص کاوش چندسطحی الگوها می‌پردازیم.

در حالت کلی روش‌های موجود برای کاوش الگوهای متوالی را می‌توان به دو دسته مبتنی بر تولید و تست کاندیدا (مانند الگوریتم‌های GSP^3 , $SPADE^2$ و ...) و تقسیم-غلبه (مانند الگوریتم $PrefixSpan^4$) تقسیم کرد. الگوریتم‌های دسته اول بر اساس الگوریتم Apriori از تکنیک تولید کاندیداها و هرس آنها بر اساس شرط حداقل پشتیبان استفاده می‌کنند، با این تفاوت که در روش SPADE از تبدیل پایگاه داده از فرمت افقی به فرمت عمودی استفاده شده و سپس کاوش الگوها انجام می‌گیرد [12]. در دسته دوم از تکنیک تقسیم فضا به حالت‌های کوچک‌تر و کاوش الگوها در فضاهای کوچک‌تر استفاده می‌کنند.

برای کاوش چندسطحی الگوها، روش توالی‌های گسترش‌یافته⁵ برای الگوریتم GSP بارزترین روش است [13]. اولین تلاش‌ها در این زمینه توسط آگراول و سریکانت [14] و هن و فو [15] انجام گرفت. پینتو و همکاران [16] و همچنین پی و همکاران [17] از روش توالی یکتا و الگوی توسعه یافته مبتنی بر $prefixSpan$ برای کاوش چندسطحی الگوها استفاده کردند. نهایت آنکه چنا و هوانگ [18] و هوانگ [19] از تئوری فازی برای حالت خاصی از کاوش چند سطحی الگوهای متوالی استفاده کرده‌اند.

چارچوب استفاده از دانش مشتریان برای

طراحی کاتالوگ‌های الکترونیکی

از آنجایی که معرفی‌نکردن الگوریتم مورد استفاده، از معایب مدل‌ها و چارچوب‌های معرفی شده تا کنون است، در این بخش در کنار معرفی چارچوب طراحی

جدول 1: متغیرهای الگوریتم DVlW

متغیرها	توضیحات
T	آرایه‌ای از یک ساختار برای ذخیره رویدادها و برجسب‌های دسته‌بندی مربوط به آنها
σ	میزان حداقل پشتیبانی
d	یک توالی از پایگاه داده (DB)
g	گرافی که توالی چندسطحی را نشان می‌دهد
I	مجموعه آیتم‌ها
C	مجموعه کاندیدها، C_k نیز کاندیدای مرحله k ام را نشان می‌دهد
L	مجموعه الگوهای مکرر، L_k نیز الگوهای مرحله k ام را نشان می‌دهد
k	شمارنده مرحله
P	مجموعه الگوهای مکرر والد، P_z نیز الگوهای مرحله k ام را نشان می‌دهد

جدول 2: توابع استفاده شده در الگوریتم DVlW

تابع	ورودی	خروجی
Make Hierarchy ()	I	T
	ساختار T را با توجه به دسته بندی تعریف شده برای آیتم‌های I پر می‌کند	
GVD()	d, T	g
	در این تابع، رخدادها رئوس گراف و توالی رخدادها یال‌ها را تشکیل می‌دهند. هر رأس مختصاتی به نام index دارد که موقعیت آن رخداد را در توالی نشان می‌دهد	
Count Support()	C_k, σ, g	Support (c_k)
	در صورتی که گراف g شامل هر یک از کاندیدهای C_k باشد، پشتیبان آن کاندیدا افزایش می‌یابد. همچنین توالی والد هر یک از کاندیدها به مجموعه P_z اضافه و پشتیبان آن والد در صورتیکه g شامل آن باشد، افزایش می‌یابد	
Prune Extra Parents()	C_k, P_z	P_z
	این تابع توالی‌های والدی را که پشتیبان آنها با مجموع پشتیبان فرزندان آنها در مجموعه C_k برابر باشد، زائد شناخته حذف می‌کند	
Candidate Generation()	L_k	C_{k+1}
	این تابع با اتصال توالی‌های مکرر، کاندیدهای مرحله بعد را برای دور بعدی اجرای الگوریتم فراهم می‌کند	

اعتبارسنجی الگوریتم پیشنهادی

برای بررسی کارایی و عملکرد الگوریتم پیشنهادی برای کاوش چندسطحی الگوهای متوالی، الگوریتم پیشنهادی با الگوریتم پایه GSP به مقایسه گذاشته شده است. بدین منظور از دو مجموعه داده ترکیبی در دو اندازه مختلف 1000 و 10 هزار آیتمی از برنامه تولید داده گروه تحقیقاتی Quest شرکت IBM مورد استفاده قرار گرفت.⁷ هر یک از این مجموعه داده‌ها ویژگی‌های خاص خود را دارند که سبب تنوع به‌کارگیری داده‌ها در مسئله می‌شوند. با توجه به نحوه توزیع الگوهای مکرر در این دو مجموعه داده، مجموعه داده اول در حداقل پشتیبان‌های 0.5، 1.0، 1.5 تا 3.0 و مجموعه داده دوم در حداقل پشتیبان‌های 0.25، 0.5، 0.75 و 1 اجرا شد. همان طور که در جدول

روش‌های معرفی شده تا به حال انجام می‌دهد. دلیل این موضوع مدیریت چندسطحی بودن توالی در مرحله‌ای جدا از تولید کاندیدا است. در پیاده‌سازی الگوریتم مسئله چندسطحی بودن، به صورت پویایی رئوس برای گراف توالی (که نشان‌دهنده رخدادها هستند) در نظر گرفته می‌شود. این الگوریتم، شامل چندین تکرار است که با توالی‌هایی به طول یک شروع شده و در هر تکرار همه الگوهای مکرر هم‌اندازه یافت می‌شوند. علاوه بر این، در هر تکرار، کاندیدها با استفاده از الگوهای مکرر به دست آمده در تکرار قبل تولید می‌شوند. الگوریتم 1 ساختار DVlW را نشان می‌دهد.

تابع GDV (Graph for Dynamic Vertices) همه حالت‌های ممکن را برای توالی d با توجه به سطوح تعریف شده برای آیتم‌ها (T) به دست می‌آورد. با استفاده از چنین تابعی کاوش در چندین سطح به صورت یک باره و هم چنین قبل و جدا از مرحله شمارش و تولید کاندیدها انجام می‌گیرد.

Algorithm 1 (The DVlW algorithm)

Data: a frequency threshold σ .

Result: The collection L of frequent sequences, k the maximal frequent length.

$L_0=0; k=1;$

$C_1 = \{ \{ \langle i \rangle \} / i \in I \}; // \text{all 1-frequent sequences}$

$T = \text{MakeHierarchy}(I)$

While $C_k \neq \emptyset$ do

For each $d \in DB$ do

$g = \text{GDV}(d, T);$

// g stands for the multilevel

graph representing of d.

$\text{CountSupport}(C_k, \sigma, g);$

$P_z = \{ p \in P_z / \text{Support}(p) > \sigma \};$

$\text{Prune Extra Parents}(C_k, P_z);$

$L_k = \{ c \in C_k / \text{Support}(c) > \sigma \};$

$C_{k+1} = \text{Candidate Generation}(L_k);$

$K=k+1;$

Return $L = \bigcup_{j=0}^k L_j;$

متغیرها و توابع مورد استفاده در الگوریتم 1، به ترتیب

در جدول (1) و (2) به نمایش گذاشته شده‌اند.

در مقایسه الگوریتم با روش توالی‌های گسترش یافته مورد استفاده در GSP، DVlW عملکرد بهتری از نظر زمان اجرا در مجموعه داده‌های مختلف نشان داد و در مورد الگوهای نهایی نیز، الگوهای مازاد تولید شده توسط GSP [1] در خروجی الگوریتم DVlW وجود نداشت.

الکترونیکی، ارتباطی با بخش‌بندی مشتریان و رفتار قبلی مشتری نداشته و برای همه یکسان در نظر گرفته می‌شود. در ادامه سعی می‌شود با پیاده‌سازی چارچوب پیشنهادی در این شرکت، استراتژی‌هایی برای طراحی کاتالوگ‌ها برای مشتریان مختلف در نظر گرفته شود.

جدول 3: نتایج اعتبارسنجی الگوریتم DVIw

مجموعه 10 هزار آیتمی					مجموعه 1000 آیتمی						
زمان DVIw (ثانیه)	زمان GSP (ثانیه)	سلاج	پشتیبان	زمان DVIw (ثانیه)	زمان GSP (ثانیه)	سلاج	پشتیبان	زمان DVIw (ثانیه)	زمان GSP (ثانیه)	سلاج	پشتیبان
370	376	1	0.25%	951	955	1	1%	951	955	1	1%
406	874	2		1014	1580	2					
451	1350	3		1097	---	3					
488	---	4		1148	---	4					
512	---	5		1206	---	5					
82	81	1	0.5%	152	150	1	1.5%	152	150	1	1.5%
98	152	2		189	287	2					
122	189	3		233	394	3					
149	248	4		265	523	4					
175	298	5		302	674	5					
50	48	1	0.75%	89	87	1	2%	89	87	1	2%
65	75	2		108	158	2					
78	104	3		130	199	3					
98	132	4		154	256	4					
118	168	5		178	305	5					
16	16	1	1%	60	58	1	2.5%	60	58	1	2.5%
28	34	2		78	89	2					
38	59	3		91	115	3					
52	81	4		109	147	4					
61	112	5		133	181	5					
				16	14	1	3%	16	14	1	3%
				32	48	2					
				57	74	3					
				69	96	4					
				93	138	5					

توجه: بخشهایی که بصورت --- نمایش داده شده است، بمعنی عدم توانایی الگوریتم در اجرا می‌باشد

پیاده‌سازی چارچوب در شرکت مورد مطالعه

داده‌های مشتریان که از راه وب سایت شرکت از تاریخ July 2006 تا May 2009 جمع‌آوری شده است، شامل 6714 تراکنش است که توسط 2487 کاربر به ثبت رسیده است. این مقدار حدود 3 درصد از فروش شرکت را شامل می‌شود. این حقیقت که فقط 3 درصد از کل فروش شرکت در این تحقیق مورد بررسی قرار می‌گیرد، به عنوان محدودیت تحقیق در نظر گرفته نمی‌شود، چرا که هدف استراتژی‌ها طراحی بهتر کاتالوگ برای مشتریانی است که در اینترنت فعال هستند و این بخش از مشتریان نیز به طور کامل در داده‌های مورد بررسی قرار دارند.

(3) مشاهده می‌شود، الگوریتم پیشنهادی DVIw نسبت به الگوریتم پایه GSP در همه اجراها برتری محسوسی دارد. در هر پشتیبانی با افزایش تعداد سطوح، تفاوت عملکرد بین دو الگوریتم نیز افزایش می‌یابد. این موضوع به دلیل افزایش طول توالی‌ها و تعداد کاندیدها در الگوریتم GSP است، در حالی که این دو مقدار در الگوریتم DVIw ثابت می‌ماند. چنانکه در جدول مشاهده می‌شود، با افزایش پشتیبان، به دلیل کاهش کلی تعداد کاندیدها، تفاوت در زمان اجرای هر دو الگوریتم کاهش می‌یابد، ولی الگوریتم DVIw همچنان عملکرد بهتری را نشان می‌دهد.

نتایج تجربی

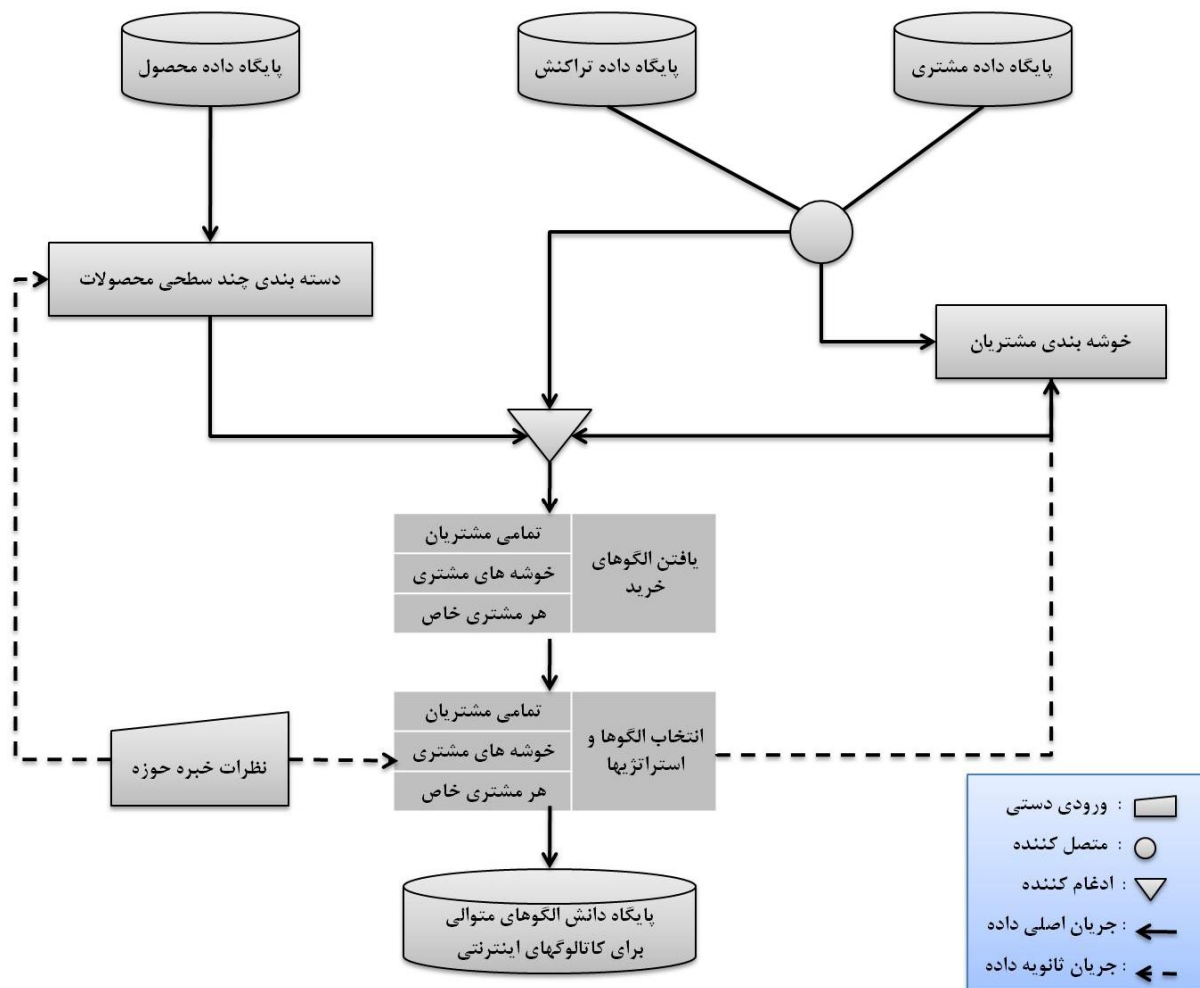
در این بخش ابتدا به معرفی شرکت مورد مطالعه می‌پردازیم، سپس نحوه پیاده‌سازی چارچوب پیشنهادی در مورد مطالعاتی تشریح می‌شود.

معرفی شرکت فروشنده تجهیزات رایانه‌ای

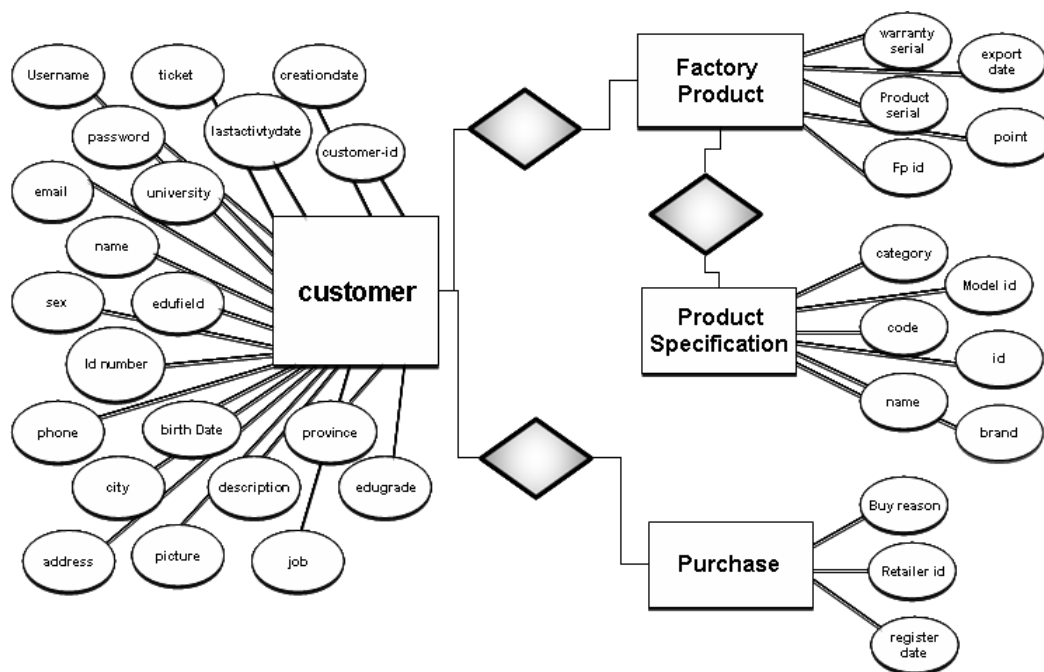
این شرکت، کار طراحی، تولید و بازاریابی تجهیزات الکترونیکی و رایانه‌ای را انجام می‌دهد و در سال 1990 کار خود را با تولید فقط یک نوع "صفحه کلید" آغاز کرده است. با دستیابی به موفقیت این محصول در بازار، شرکت گسترش افقی خود را برای دستیابی به بازار هدف بزرگ‌تر و تولید محصولات و تجهیزات رایانه‌ای بیشتر آغاز کرد. در حال حاضر، کمپانی بیش از صد نوع محصول در دسته‌بندی‌های مختلف مانند رایانه‌های قابل حمل (لپ تاپ)، تلویزیون LCD، نمایشگرهای LCD، دستگاه‌های چند رسانه‌ای، ضبط و پخش DVD و VCD، دستگاه‌های چندرسانه‌ای و بیسیم، موشواره‌های نوری و ... دارد.

شرکت با این تنوع محصولات به دنبال راه‌هایی برای گسترش بازار هدف خود با استفاده از فناوری‌های جدید نظیر فناوری اطلاعات است. در اولین گام، برای دستیابی به اطلاعات مشتریان و رفتار خرید آنها، یک وب سایت با هدف ایجاد ارتباط با مشتری، توسعه برنامه‌های وفاداری و نظارت بر توزیع‌کنندگان ایجاد شد. شکل (2) قسمتی از ساختار پایگاه داده این وب سایت را نشان می‌دهد.

ارسال پست الکترونیکی برای اعضا که شامل اطلاعات محصولات جدید بود، اولین ابزار تبلیغاتی شرکت با استفاده از اطلاعات مشتریان است. اما محتوای این پست



شکل 1: چارچوب طراحی کاتالوگهای الکترونیکی بر اساس دانش مشتری



شکل 2: بخشی از مدل E-R پایگاه داده کمپانی مورد مطالعه

خوشه‌بندی مشتریان

خوشه‌بندی روی پایگاه داده مشتریان با استفاده از الگوریتم K-means و نرم‌افزار SPSS Clementine انجام گرفت. با توجه به نظر خبره حوزه و بر اساس اطلاعات موجود در پایگاه داده، ویژگی‌های سال تولد و میزان تحصیلات برای تشکیل خوشه‌ها انتخاب شدند. برای اجرای الگوریتم k-means، مقدار k در یک محدوده قابل قبول $k \in [3, 9]$ اجرا شد. پس از ارزیابی خوشه‌ها از نظر میزان خطا و میزان همگنی آنها (نبود خوشه‌های یک عضوی یا دو-سه عضوی) توسط خبره حوزه، عدد 4 به عنوان بهترین حالت برای تعداد خوشه‌ها در نظر گرفته شد. اجرای روش دو مرحله‌ای ارائه شده توسط نرم‌افزار SPSS نیز بر درستی این انتخاب صحت گذاشت. جدول (4) نماینده هر خوشه را نشان می‌دهد.

جدول 4: خوشه‌های مشتریان در کمپانی مورد مطالعه

شماره خوشه	ویژگی‌های نماینده خوشه		تعداد مشتریان خوشه
	میزان تحصیلات	سال تولد	
1	فوق دیپلم	1965	251
2	لیسانس	1973	473
3	لیسانس	1985	1121
4	دیپلم	1991	642

دسته‌بندی محصولات

در این مرحله با توجه به نظر خبره حوزه برای محصولات شرکت دسته‌بندی تا حداکثر سه سطح در نظر گرفته شد. جدول (5) چگونگی دسته‌بندی را برای محصولات شرکت نشان می‌دهد. این دسته‌بندی در هر سطحی می‌تواند برای همه یا برخی از موارد انجام شود و مواردی که در سطح L برچسبی نمی‌گیرند، از برچسب پیش‌فرض مورد برای الگوهای متوالی آن سطح استفاده می‌کنند.

همان‌طور که در جدول (5) آمده است، در دسته‌بندی سطح 1، محصولات یک خانواده در یک دسته‌بندی قرار گرفته‌اند (برای مثال انواع موشواره‌ها در دسته‌بندی با برچسب 6). در دسته‌بندی سطح 2 وسایل ورودی و خروجی با قیمت کم به ترتیب با برچسب‌های A و B نشان داده شده‌اند و در سطح 3 نیز، موارد بر اساس حاشیه سود برای شرکت در سه دسته‌بندی با حاشیه سود کم، متوسط و زیاد قرار گرفته‌اند.

یافتن الگوهای متوالی

همان‌طور که چارچوب طراحی کاتالوگ‌های الکترونیکی نشان می‌دهد، الگوها باید در سه سطح همه مشتریان، خوشه‌های مشتریان و هر مشتری کاوش شوند. از آنجا که به طور متوسط در توالی مربوط به هر مشتری کمپانی 2/7 رویداد وجود دارد، یافتن الگوها در سطح هر مشتری نتیجه معناداری را در برنخواهد داشت، بنابراین از آن صرف‌نظر می‌کنیم. حداقل پشتیبان نیز 10 درصد در نظر گرفته شده است.

با اعمال الگوریتم DVIw روی داده‌های همه مشتریان در سه سطح تعریف شده برای محصولات در مرحله قبل، الگوهای زیر به دست می‌آید (جدول 6).

اکنون الگوریتم کاوش الگوهای متوالی را برای چهار خوشه در نظر گرفته شده برای مشتریان اعمال می‌کنیم. (جدول 7 تا 10 الگوهای یافت شده برای هر یک از خوشه‌ها را نشان می‌دهد.)

دانش مشتریان برای طراحی کاتالوگ‌های الکترونیکی

پس از دستیابی به الگوهای متوالی، می‌توان با کمک خبره حوزه آنها را بررسی و استراتژی‌هایی را بر اساس آنها تدوین کرد. با بررسی الگوهای به دست آمده در مراحل قبل، دانش زیر قابل بیان است.

در سطح همه مشتریان:

1. در دو بار (یا چندین بار) خرید متوالی تکرار در خرید موشواره و صفحه کلید دیده می‌شود.
2. در دو بار خرید، با خرید موشواره یا صفحه کلید، LCD و اسپیکر نیز خریده شده است.
3. تکرار در خرید وسایل ورودی ارزان قیمت (با برچسب A) وجود دارد.
4. بعد از خرید در گروه وسایل ورودی ارزان قیمت (با برچسب A)، خریدی در گروه وسایل خروجی ارزان قیمت (با برچسب B) یا خرید LCD رخ داده است.
5. تکرار خرید در اجناس با حاشیه سود کم (با برچسب L) وجود دارد. همچنین در دو بار خرید، بعد از خرید وسیله‌ای در گروه حاشیه سود کم، وسیله‌ای در گروه حاشیه سود بالا (با برچسب H) و بر عکس خریداری شده است.

در مورد خوشه شماره 3:

1. در توالی‌های خرید موارد موشواره و صفحه کلید به ترتیب پشت سر یکدیگر قرار دارند.
2. تکرار در خرید وسایل ورودی ارزان قیمت (با برچسب A) وجود دارد.
3. تکرار خرید در اجناس با حاشیه سود کم (با برچسب L) وجود دارد.

در مورد خوشه شماره 4:

1. بعد از خرید صفحه کلید یا موشواره، Fpack یا LCD و case خریداری شده است.
2. بعد از خرید وسایل ورودی ارزان قیمت (با برچسب A)، خرید case یا Fpack یا LCD انجام شده است.
3. تکرار خرید در اجناس با حاشیه سود کم (با برچسب L) وجود دارد. همچنین در دو بار خرید، بعد از خرید وسیله‌ای در گروه حاشیه سود کم، وسیله‌ای در گروه حاشیه سود بالا (با برچسب H) و بر عکس انجام شده است.

با توجه به دانش به دست آمده در سطح خوشه‌ها و همه مشتریان، شرکت استراتژی‌هایی را برای طراحی کاتالوگ‌های الکترونیکی در نظر گرفت (جدول 11). در مورد مشتریان به طور کلی، فروش محصولات با حاشیه سود متوسط در قالب پیشنهادهای فروش جانبی در کنار محصولات پرفروش، می‌تواند مورد توجه قرار گیرد. همچنین از این روش می‌توان برای محصولات دسته B که نسبت به محصولات در دسته A فروش کمتری دارند، استفاده کرد. با توجه به فروش بالای صفحه کلید و موشواره، پیشنهاد بسته‌های آماده رایانه‌ای برای مشتریان بالقوه نیز استراتژی دیگری برای مشتریان در سطح عام است. در مورد خوشه‌های مشتریان 1 و 2 نیز تبلیغ بسته‌های محصولات و در مورد خوشه 3، تشویق مشتریان برای خرید اجناس با قیمت بالاتر به وسیله تخفیف‌ها از استراتژی‌های در پیش گرفته شده است. در خوشه 4 با توجه به فروش متنوع‌تر محصولات تبلیغ بسته‌های محصولات کامل‌تر پیشنهاد می‌شود.

در سطح خوشه‌های مشتریان، در مورد خوشه شماره 1:

1. در توالی‌های خرید آیتم‌های موشواره و صفحه کلید به ترتیب پشت سر یکدیگر قرار دارند. همچنین خرید LCD نیز در توالی‌های به طول 2، بعد از خرید موشواره یا صفحه کلید و در توالی به طول 3، قبل از خرید این دو روی داده است.
2. تکرار در خرید وسایل ورودی ارزان قیمت (با برچسب A) وجود دارد. همچنین LCD و وسایل گروه A به طور متوالی خریداری شده‌اند. بعد از خرید وسیله در گروه A، وسیله‌ای در برچسب B خریداری شده است.
3. تکرار خرید در اجناس با حاشیه سود کم (با برچسب L) وجود دارد. همچنین در دو بار خرید، بعد از خرید وسیله‌ای در گروه حاشیه سود کم، وسیله‌ای در گروه حاشیه سود بالا (با برچسب H) و بر عکس خریداری شده است. در این خوشه همچنین خرید در گروه با حاشیه سود متوسط (برچسب M) بعد از خرید در گروه حاشیه سود کم روی داده است.

در مورد خوشه شماره 2:

1. در توالی‌های خرید موارد موشواره و صفحه کلید به ترتیب پشت سر یکدیگر قرار دارند. همچنین توالی‌ها خرید LCD بعد از صفحه کلید را نشان می‌دهند.
2. تکرار در خرید وسایل ورودی ارزان قیمت (با برچسب A) وجود دارد. همچنین LCD و وسایل گروه A به طور متوالی خریداری شده‌اند. بعد از خرید وسیله در گروه A، وسیله‌ای با برچسب B خریداری شده است و بر عکس.
3. تکرار خرید در اجناس با حاشیه سود کم (با برچسب L) وجود دارد. همچنین در دو بار خرید، بعد از خرید وسیله‌ای در گروه حاشیه سود کم، وسیله‌ای در گروه حاشیه سود بالا (با برچسب H) و بر عکس خریداری شده است.

جدول 5: دسته‌بندی محصولات در شرکت مورد مطالعه

برچسب در سطح 3	برچسب در سطح 2	برچسب در سطح 1	محصولات
H		1	Notebook
H		2	pack (2..8 devices)
H		3	LCD(TV-monitor)
M	A	4	Desktop solution(mouse and keyboard)
L	A	5	Keyboard
L	A	6	Mouse
L	B	7	Speaker
L	B	8	Game solution
L	B	9	Headset
L		10	Case
M		11	Media player (mp3-mp4-...)
M		12	Flash driver
L	A	13	Webcam
M		14	Memory
M		15	Optical driver
M		16	Power supply
M		17	Accessories

جدول 6: الگوها برای همه مشتریان

طول الگو	سطح 1	سطح 2	سطح 3
2	<6 3> <6 5> <6 7> <5 6> <3 5>	<A B> <A 3> <3 A> <3 3> <A A>	<L L> <H H> <L H> <H L>
3	<5 6 5> <6 6 5> <6 (5,6)> <5 (5,6)>	<3 A A> <A A A>	<L L L>
4	<6 (5,6) 5>	<A A A A>	<L L L L>

جدول 7: الگوها برای خوشه 1

طول الگو	سطح 1	سطح 2	سطح 3
2	<6 3> <5 7> <5 3> <5 6> <6 5>	<A A> <A 3> <3 A> <A B>	<L L> <L M> <L H> <H L>
3	<3 5 6>	<A A A>	<L L L>
4	<6 (5,6) 6>	<A A A A>	<L L L L>
5	<6 (5,6)(5,6)>	<A A A A A>	<L L L L L>

جدول 8: الگوها برای خوشه 2

طول الگو	سطح 1	سطح 2	سطح 3
2	<6 3> <5 6> <6 5>	<A A> <A 3> <B A> <A B>	<L L> <L H> <H L>
3	<6 (5,6)>	<A A A> <A (B,A)>	<L L L>
4	---	<A A A A>	<L L L L>
5	---	<A A A A A>	---

جدول 9: الگوها برای خوشه 3

طول الگو	سطح 1	سطح 2	سطح 3
2	<5 6> <6 5>	<A A>	<L L>
3	---	<A A A>	<L L L>
4	---	<A A A A>	<L L L L>
5	---	---	---

جدول 10: الگوها برای خوشه 4

طول الگو	سطح 1	سطح 2	سطح 3
2	<6 10> <5 2> <6 3> <10 3> <5 6> <3 5>	<A 10> <A 2> <A 3> <3 A>	<L L> <L H> <H L>
3	---	---	<L L L>
4	---	---	<L L L L>
5	---	---	---

حالت کاوش چندسطحی الگوها با توجه به خوشه‌بندی مشتریان، مجموع دانش‌های مراحل قبل حاصل می‌شود. در مورد مطالعاتی نیز، مدیریت شرکت الگوهای نهایی از چارچوب به دست آمده را کامل‌تر از حالات قبلی و مورد قبول دانست.

نتیجه‌گیری







در این تحقیق، چارچوبی برای استفاده از دانش مشتریان در طراحی کاتالوگ‌های الکترونیکی معرفی شد که دانش مشتریان را با استفاده از سطوح مختلف دسته‌بندی محصولات و در سه گروه همه مشتریان، خوشه‌های مشتریان و هر مشتری به دست می‌آورد. با پیاده‌سازی این چارچوب در یک شرکت فروشنده تجهیزات رایانه‌ای، مشاهده شد که در مقایسه با چارچوب‌های ارائه شده تا به حال دانش کامل‌تری از پایگاه داده‌های شرکت استخراج شد. با بهره‌گیری از این دانش و نظر خبره حوزه، استراتژی‌هایی برای طراحی کاتالوگ‌های الکترونیکی در این شرکت تهیه شد که بر خلاف حالت اولیه سبب سفارشی‌سازی در کاتالوگ‌ها شد.

در نظر گرفتن استراتژی‌های قیمت‌گذاری در چارچوب طراحی کاتالوگ‌های الکترونیکی برای کامل‌تر کردن الگوهای به دست آمده و ایجاد بخشی برای امتیازدهی و ارزیابی نهایی الگوهای به دست آمده از داده‌کاوی از مباحثی است که می‌تواند در مطالعات آتی مورد توجه قرار گیرد.

بررسی دانش به دست آمده

روشن است که کاوش الگوهای متوالی در چندین لایه و سطوح همه مشتریان و خوشه‌های مشتریان دانش کامل‌تری را نسبت به سایر حالت‌ها به دست می‌دهد. در جدول (12) دانش در حالتی که الگوها در یک لایه یا بدون استفاده از خوشه‌ها کاوش شوند، در کنار دانش به دست آمده از طریق چارچوب طراحی کاتالوگ‌های اینترنتی معرفی شده، نشان داده شده است. در کاوش الگوها در یک سطح و برای همه مشتریان فقط ارتباط بین خرید موشواره و صفحه کلید و LCD و اسپیکر کشف شده است و الگوهایی که با کاوش چندسطحی الگوهای متوالی و با توجه به خوشه‌بندی مشتریان به دست می‌آیند در نظر گرفته نمی‌شوند. در حالتی که تنها به خوشه‌بندی مشتریان توجه شود، در کنار ارتباط بین خرید موشواره و صفحه کلید و LCD و اسپیکر، الگوهایی که نشان می‌دهد در توالی‌های خرید خوشه افراد کم سن نسبت به سایر خوشه‌ها تفاوت‌هایی وجود دارد نیز کشف می‌شود، اما این الگوها در سطح یک متوقف می‌شوند. در حالتی که بدون توجه به خوشه‌بندی مشتریان تنها دسته‌بندی محصولات در نظر گرفته شود، وجود توالی در خرید وسایل ورودی ارزان قیمت و تجهیزات با حاشیه سود کم از مواردی است که علاوه بر دانش حالت اول، کشف می‌شود. در نهایت در

جدول 11: نمونه ای از استراتژی‌های طراحی کاتالوگ‌های الکترونیکی در مورد مطالعاتی

مثال	نمونه استراتژی	سطح مشتریان
	اولویت در تبلیغ بسته‌های محصولات	خوشه 1
	اولویت در تبلیغ بسته‌های محصولات	خوشه 2
 + تخفیف	تشویق مشتریان برای خرید اجناس با قیمت بالا	خوشه 3
	اولویت در تبلیغ بسته‌های محصولات بزرگ شامل LCD و Case	خوشه 4
 + تخفیف	قرار دادن محصولات با حاشیه سود متوسط در کنار محصولات با حاشیه سود کم یا زیاد در قالب پیشنهادات فروش جانبی	تمام مشتریان
 + تخفیف	قرار دادن وسایل خروجی ارزان قیمت در کنار وسایل ورودی در قالب پیشنهادات فروش جانبی	
 پیشنهاد به جای یا	تبلیغ بسته‌های آماده رایانه ای (Desktop Solution) برای مشتریان بالقوه در قالب پیشنهاد فروش بالاسری	

جدول 12: دانش به دست آمده در چارچوب پیشنهادی در مقایسه با سایر حالتها

ارتباط بین خرید موشواره و صفحه کلید و LCD و اسپیکر در توالی‌ها نشان داده می‌شود. وجود توالی در خرید وسایل ورودی ارزان قیمت، خرید وسایل خروجی ارزان قیمت بعد از وسایل ورودی ارزان قیمت، خرید LCD قبل و بعد از وسایل ورودی ارزان قیمت.	کاوش چندسطحی الگوها- با توجه به خوشه بندی مشتریان
وجود توالی در خرید وسایل با حاشیه سود کم، خرید وسایل با حاشیه سود بالا قبل و بعد از وسایل با حاشیه سود کم.. تفاوت توالی‌های خرید موجود در خوشه افراد کم سن (وجود case وFPack در توالی‌ها در کنار موشواره و صفحه کلید و LCD) با سایر خوشه‌ها سایر الگوهایی که از کاوش توالی‌ها در چندین سطح و با خوشه بندی مشتریان بدست آمده است.	کاوش چندسطحی الگوها - تمام مشتریان
ارتباط بین خرید موشواره و صفحه کلید و LCD و اسپیکر در توالی‌ها نشان داده می‌شود. وجود توالی در خرید وسایل ورودی ارزان قیمت، خرید وسایل خروجی ارزان قیمت بعد از وسایل ورودی ارزان قیمت، خرید LCD قبل و بعد از وسایل ورودی ارزان قیمت.	کاوش الگوها در یک سطح اما با توجه به خوشه بندی مشتریان [10]
ارتباط بین خرید موشواره و صفحه کلید و LCD و اسپیکر در توالی‌ها نشان داده می‌شود. تفاوت توالی‌های خرید موجود در خوشه افراد کم سن (وجود case وFPack در توالی‌ها در کنار موشواره و صفحه کلید و LCD) با سایر خوشه‌ها	کاوش الگوها در یک سطح - تمام مشتریان [9, 8]
تنها ارتباط بین خرید در توالی‌ها نشان داده می‌شود.	

مراجع

- 1- Agrawal, R. and Srikant, R. (1996). "Mining sequential patterns: generalizations and performance improvements." *Proc., 5th Int. Conf. on Extending Database Technology*, PP. 3-17.
- 2- Gibbert, M., Leibold, M. and Probst, G. (2002). "Five styles of customer knowledge management, and how smart companies use them to create value." *European Management Journal*, Vol. 20, No. 5, PP. 459-469.
- 3- Gebert, H., Geib, M., Kolbe, L.M. and Brenner, W. (2003). "Knowledge-enabled customer relationship management." *Journal of Knowledge Management*, Vol. 7, No. 5, PP.107-123.
- 4- Salomann, H., Dous, M., Kolbe, L. and Brenner, W. (2005) "Rejuvenating customer management: How to make knowledge for, from and about customers work." *European Management Journal*, Vol. 23, No. 4, PP. 392-403.
- 5- Paquette, S. (2008), "Knowledge management systems and customer knowledge use in organizations." Doctoral Dissertation, Faculty of Information Studies, University of Toronto.
- 6- Lopez-Nicolas, C. and Molina-Castillo, F. (2008). "Customer knowledge management and e-commerce: The role of customer perceived risk." *International Journal of Information Management*, Vol. 28 , PP. 102-113.
- 7- Su, C-T., Chen, Y-H. and Sha, D.Y. (2006). "Linking innovative product development with customer knowledge:a data-mining approach." *Technovation*, Vol. 26, PP.784-795.
- 8- Liao, S.H. and Chen, Y.J. (2004). "Mining customer knowledge for electronic catalog marketing." *Expert Systems with Applications*, Vol. 27, PP.521-532.
- 9- Lin, C. and Hong, C. (2008). "Using customer knowledge in designing electronic catalog." *Expert Systems with Applications*, Vol. 34, PP. 119-127.
- 10- Changchien, S. W., Lee, C.-F., and Hsu, Y.-J. (2004). "On-line personalized sales promotion in electronic commerce." *Expert Systems with Applications*, Vol. 27, No. 1, PP. 35-52.
- 11- Prinzie, A. and Poel D. (2006). "Investigating purchasing patterns for financial services using markov, MTD and MTDg Models." *European Journal of Operational Research*, Vol. 170, No .3, PP. 710-734.
- 12- Zaki, M.J. (2001). "Spade: An efficient algorithm for mining frequent sequences." *Machine Learning*, Vol. 42, No. 1-2, PP. 31-60.

- 13- Agrawal, R., and Srikant, R. (1995). "Mining generalized association rules." *Proc., Int. Conf. on Very Large Data Bases*, Zurich, PP. 407-419.
- 14- Agrawal, R., and Srikant, R. (1995). "Mining sequential patterns." *Proc., 11th Conference on Data Engineering*, Taipei, Taiwan.
- 15- Han, J., and Fu, Y. (1999). "Mining multiple-level association rules in large databases." *IEEE Transaction on Knowledge and Data Engineering*, vol.11, No. 5, PP. 1-8.
- 16- Pinto, H., Han, J., Pei, J., Wang, K., Chen, Q. and Dayal. U. (2001). "Multi-dimensional sequential pattern mining." *Proc., Int. Conf. Information and Knowledge Management (CIKM'01)*, Atlanta, GA, PP. 81-88.
- 17- Pei J, Han J. et al. (2001). "PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth." *Proc., Int. Conf. on Data Engineering*, PP. 215-226.
- 18- Chena, Y.L. and Huang, T. C.K.(2008). "A novel knowledge discovering model for mining fuzzy multi-level sequential patterns in sequence databases". *Data & Knowledge Engineering*, Vol. 66, No. 3, PP. 349-367.
- 19- Huang, T. C.K. (2009). "Developing an efficient knowledge discovering model for mining fuzzy multi-level sequential patterns in sequence databases." *Fuzzy Sets and Systems*, Vol. 160, No. 23, PP. 3359-3381.

واژه‌های انگلیسی به ترتیب استفاده در متن

- 1- eCKM: Electronic Customer Knowledge Management
- 2- SPADE: Sequential PAttern Discovery Using Equivalent Classes
- 3- GSP: Generalized Sequential Pattern Algorithm
- 4- Prefix-Projected Sequential Pattern Growth
- 5- Extended Sequences
- 6- Dynamic Vertices Level Wise
- 7- http://www.almaden.ibm.com/cs/projects/iis/hdb/Projects/data_mining/datasets/syndata.html