



دانشگاه صنعتی امیرکبیر

[پلی تکنیک تهران]

دانشکده مهندسی صنایع

پایان نامه کارشناسی ارشد

گرایش مدیریت سیستم و بهره وری

به کارگیری تکنیک های داده کاوی در کشف دانش فروش با در نظر گرفتن
مجاورت کالاها در قفسه ها بر اساس معیار قیمت

نگارش

مریم نفری

استاد راهنما

دکتر جمال شهبابی



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

بسمه تعالی

تاریخ:
شماره:

فرم اطلاعات پایان نامه کارشناسی - ارشد و دکترا

معاونت پژوهشی
فرم پروژه تحصیلات تکمیلی 7

مشخصات دانشجو:

نام و نام خانوادگی: مریم نفری
شماره دانشجویی: 85125004
دانشکده: مهندسی صنایع رشته تحصیلی: مهندسی صنایع گروه: مدیریت سیستم و بهره وری
دانشجوی آزاد بورسیه معادل

مشخصات استاد راهنما:

نام و نام خانوادگی: جمال شهرابی
نام و نام خانوادگی:
درجه و رتبه: دکترا
درجه و رتبه:

مشخصات استاد مشاور:

نام و نام خانوادگی:
نام و نام خانوادگی:
درجه و رتبه:
درجه و رتبه:

عنوان پایان نامه به فارسی: به کارگیری تکنیک های داده کاوی در کشف دانش فروش با در نظر گرفتن مجاورت کالاها در قفسه ها بر اساس معیار قیمت
عنوان پایان نامه به انگلیسی:

Applying data mining techniques for retail knowledge discovery with consideration of shelf-space adjacency based on price attribute

نوع پروژه: کارشناسی ارشد دکترا سال تحصیلی: 85-87
کاربردی بنیادی توسعه ای نظری

تاریخ شروع: 85 تاریخ خاتمه: 87 تعداد واحد: 32 سازمان تأمین کننده اعتبار:

واژه های کلیدی به فارسی: داده کاوی، قوانین وابستگی زمان محور، تخفیف، مدیریت فضای قفسات، انتخاب کالاها

واژه های کلیدی به انگلیسی: Data mining, Multiple-level temporal association rules, Discount, Shelf-space management, Product assortment

مشخصات ظاهری	تعداد صفحات	تصویر <input checked="" type="radio"/> جدول <input checked="" type="radio"/> نمودار <input checked="" type="radio"/> نقشه <input type="radio"/> واژه نامه <input type="radio"/>	تعداد مراجع	تعداد صفحات ضمیمه
	122		97	22
زبان متن	فارسی <input checked="" type="radio"/> انگلیسی <input type="radio"/>	چکیده	فارسی <input checked="" type="radio"/> انگلیسی <input type="radio"/>	
یادداشت				

نظرها و پیشنهادهای به منظور بهبود فعالیت های پژوهشی دانشگاه

استاد:

دانشجو:

امضاء استاد راهنما: تاریخ:

تقدیم به

روح پاک پدر عزیزم که تمام انگیزه من برای راهی شدن در این مسیر مقدس بود، تا روزی جای پای او بگذارم و بزرگترین آرزویش را جامه عمل بپوشانم.

تقدیم به

مادر گرانقدرم که خالصانه قدم به قدم در راه پیشرفت من در کنارم بود و در سخت ترین روزها از هیچ کمکی در حق من دریغ نکرد.

تشکر و قدردانی

حمد و سپاس مخصوص خداست که پروردگار جهانیان است. اکنون که در سایه الطاف و عنایت خداوند موفق به نگارش این پایان نامه شده ام، بر خود لازم می دانم از استاد گرامی جناب آقای دکتر شهرابی به خاطر رهنمودهای ارزنده و مساعدت های بی دریغ ایشان در طی این تحقیق، کمال تشکر و قدردانی را داشته باشم و از خداوند متعال سلامتی و توفیقشان را خواستارم.

چکیده

با افزایش رقابت در صنعت فروش، فروشندگان می‌کوشند عملکردشان را در راستای افزایش سودآوری و کارایی فروشگاه بهبود بخشند. از این رو، اجباراً مراکز فروش بیشتر از گذشته به مسائل زیربنایی فروشگاه همچون محرک‌ها جهت فروش کالاها، انتخاب کالاها برای نمایش و فضای نمایش کالاها توجه می‌کنند. تا کنون محرک تخفیف کالاها و اثرات آن بر روی فروش کمتر به صورت آکادمیک مورد بررسی و تحلیل قرار گرفته است. الگوریتم‌های محدود مطرح شده در این زمینه، خاصیت زمان محور بودن تخفیف کالاها و موقتی بودن آن را نادیده گرفته‌اند. در فاز اول این پایان‌نامه با استفاده از تکنیک‌های داده‌کاوی، الگوریتمی ارائه شده است که ارتباط فروش کالاها و تخفیفاتشان را کاوش کرده و از نتایج آن در فاز دوم پایان‌نامه برای انتخاب کالاها و تخصیص بهترین تخفیف به آن‌ها استفاده شده است. مدل ارائه شده برای انتخاب کالاها، یک مدل بهینه‌سازی است و بر پایه برنامه‌ریزی خطی صفر-یک بیان شده است. در این رویکرد، اثرات ضربدری فروش کالاها با تخفیفات مختلف لحاظ گردیده است و در نهایت هدف مدل بهینه‌سازی، ماکزیمم کردن سود کل حاصل از فروش منفرد کالا و فروش ضربدری کالاها است. مدل مطرح شده با استفاده از نرم افزار LINGO 9.0 حل شده است.

در فاز سوم پایان‌نامه با استفاده از نتایج حاصل از الگوریتم مطرح شده در فاز اول پایان‌نامه و قوانین وابستگی میان کالاها با تخفیفات مختلف، رویکردی برای چیدمان کالاها در قفسه‌های فروشگاه ارائه گردیده است. در آخر، جهت بررسی کارایی مدل، مطالعه تجربی روی داده‌های شبیه‌سازی شده صورت گرفته است. زبان برنامه‌ریزی مورد استفاده برای پیاده‌سازی الگوریتم Visual Basic 2000 می‌باشد و پایگاه داده Microsoft SQL Server 2000 برای ذخیره‌سازی اطلاعات و بهره‌برداری از آن‌ها بکاربرده شده است.

واژگان کلیدی

داده‌کاوی، قوانین وابستگی چند سطحی زمان محور، تخفیف، مدیریت فضای قفسات، انتخاب کالاها

فهرست مطالب

1) کلیات تحقیق و پیشینه

تاریخی.....2

1-1

مقدمه.....

2

2-1 هدف از اجرای پایان نامه و ضرورت

تحقیق.....3

3-1 روش تحقیق و نتایج کسب

شده.....5

1-3-1 داده

کاوی.....5

1-1-3-1 مقایسه داده کاوی و علم

آمار.....6

2-1-3-1 اهداف داده

کاوی.....7

3-1-3-1 انواع داده

کاوی.....8

4-1-3-1 کاربرد های داده

کاوی.....14

5-1-3-1 فواید و نقش داده کاوی در فعالیت شرکت

ها.....15

6-1-3-1 محدودیت های داده

کاوی.....16

7-1-3-1 نرم افزار.....17

2-3-1 کشف دانش از پایگاه های

داده.....18

3-3-1 داده کاوی در تحلیل سبد

بازار.....20

1-3-3-1 اهمیت تحلیل سبد

بازار.....22

4-3-1 قوانین

وابستگی.....26

1-4-3-1 تعریف مسأله قوانین

وابستگی.....27

5-3-1 انتخاب کالاها و تخصیص کالاها به قفسه

ها.....30

2) معرفی الگوریتم جدیدی برای کشف قوانین وابستگی چند سطحی با در نظر

گرفتن نرخ

.....	تخفیف		
			36
			1-2
.....	مقدمه		
			37
تعارف			2-2
38	مسأله		
Apriori-	الگوریتم		3-2
42.....		Td	
	4-2 قوانین وابستگی چند سطحی و الگوریتم Apriori-		
		TdMI	50
چند	وابستگی	قوانین	1-4-2
		سطحی	51.....
Apriori-	الگوریتم		2-4-2
52.....		TdMI1	
Apriori-	الگوریتم		3-4-2
58.....		TdMI2	
	3 انتخاب کالاهای سود آور و چیدمان آن ها در		
	قفسات		59
			1-3
.....	مقدمه		
			60

2-3 انتخاب کالاهاى سود آور.....61

1-2-3 برآورد سود ناخالص الكوهای مکرر.....61

2-2-3 انتخاب کالاها.....64

3-3 تخصیص کالاها به قفسه

ها.....68

4) آزمون و اعتبارسنجی

مدل.....74

1-4 اعتبارسنجی

مدل.....75

1-1-4 تولید داده.....75

2-4 اعتبار سنجی الگوریتم کشف قوانین وابستگی با اطلاعات تخفیف کالاها.....81

1-2-4 تغییر تعداد تراکنش های پایگاه داده D.....82

2-2-4 بررسی حمایت محلی ls.....84

3-2-4 درصد کالاها در کلاس های مختلف.....85

4-2-4 میانگین تعداد تغییرات در کلاس های MC ، HC و LC 87

3-4 اعتبار سنجی مدل انتخاب کالاها 88

4-4 اعتبار سنجی رویکرد چیدمان کالاها 90

(5) تحقیقات آینده 94

(6) مراجع 95

(7) ضمائم 101

7-1 شرح توابع دو الگوریتم $Apriori$ - و $Apriori-TdM1$ 101
TdM2

فهرست اشکال

1-1 ارتباط داده کاوی و فرآیند کشف

دانش.....19

1-2 الگوریتم Apriori-

Td.....45

2-2 تابع کشف کاندیداهای محلی 1-d-الگو

(LC₁).....45

3-2 تابع کشف d-k-الگوهای

مکرر.....46

4-2 تابع کشف کاندیداهای محلی d-k-

الگو.....47

5-2 مثالی از یک سلسله مراتبی

مفهومی.....51

6-2 الگوریتم Apriori-

TdMI.....54

1-4 تعداد الگوهای مکرر در سطح

سوم.....83

2-4 تعداد الگوهای مکرر در سطح

اول.....84

3-4 زمان اجرا برای ترکیب های PHC/PMC/

PLC.....86

4-4 تعداد الگوها برای ترکیب *PHC/PMC/*

86.....*PLC*

5-4 زمان اجرا برای ترکیب های *MC, HC* و

87.....*LC*

6-4 تعداد الگوها برای ترکیب های *MC, HC* و

87.....*LC*

7-4 تأثیر روی سودآوری با جایگزینی

89.....کالاها

8-4 تعداد دسته های موجود در بزرگترین دسته

91.....مجازی

9-4 تعداد دسته های

91.....مجازی

10-4 تعداد زیردسته های موجود در بزرگترین زیردسته

92.....مجازی

11-4 تعداد زیردسته های

92.....مجازی

12-4 تعداد آیتم های موجود در بزرگترین آیتم

93.....مجازی

13-4 تعداد آیتم های

93.....مجازی

فهرست جداول

	1-1 خلاصه مرور
35.....	ادبیات.....
	1-2 جدول تراکنشی
39.....	D
	2-2 لیست تخفیف
40.....	P
	1 3-2-الگوهای مکرر
47.....	خارجی.....
	4-2 تولید LC_1 از روی
48.....	GL_1
	5-2 تولید LL_1 از روی
49.....	LC_1
	6-2 تولید LC_2 از روی
49.....	LL_1
	7-2 تولید $d-2$ -الگوهای
50.....	مکرر.....
	8-2 پایگاه داده کد شده
53.....	$\Gamma[1]$
	9-2 جدول تخفیف
55.....	GP
	10-2 پایگاه داده
56.....	$\Gamma[2]$
	11-2 نتایج حاصل از درخت
56.....	تخفیف.....

12-2 تولید LC_2^1 از روی	
56.....	LL_1^1
13-2 تولید LL_2^3 از روی	
57.....	LC_2^3
1-3 پارامترهای مسأله برآورد سود d -الگوهای	
62.....	مکرر
2-3 پارامترهای مسأله انتخاب کالاها با تخفیف	
66.....	مناسب
3-3 متغیرهای تصمیم مسأله انتخاب کالاها با تخفیف	
66.....	مناسب
3-4 پارامترهای مسأله تخصیص کالاها به قفسه	
70.....	ها
1-4 پارامترهای استفاده شده در تولید پایگاه داده	
76.....	تراکنشی
2-4 پارامترهای استفاده شده در تولید لیست	
79.....	تخفیف
3-4 پارامترهای استفاده شده در کامل کردن پایگاه داده	
80.....	تراکنشی
4-4 پارامترهای استفاده شده در تولید	
81.....	قیمت
5-4 بازه های	
81.....	تخفیف
6-4 عملکرد نسبی سه الگوریتم $ML-T2L1$ ، $Apriori-TdM11$ و $Apriori$ -	
83.....	$TdM12$

7-4 مقایسه زمان اجرای الگوریتم Apriori-TdM11 با تغییر σ_{l_i} و σ_{g_i}85

فصل اول

کلیات تحقیق و پیشینه تاریخی

1) کلیات تحقیق و پیشینه تاریخی

1-1 مقدمه

با رشد فزاینده حجم داده‌های ذخیره شده در فایل‌ها، پایگاه‌های داده و دیگر انبارها، و همچنین نیاز روز افزون جهت دستیابی به اطلاعات ارزشمند و دانش از این داده‌های خام، داده کاوی به عنوان روشی مهم و پرکاربرد برای استخراج اطلاعات و ارضاء این نیاز مطرح شده است. در واقع داده کاوی¹ بخشی از فرایند استخراج دانش² است که در آن الگوهای مفید و مورد علاقه در پایگاه داده‌ها جستجو می‌شوند. در حال حاضر در اکثر کشورها، سازمان‌ها، شرکت‌ها، فروشگاه‌ها و . . . برای امور بازرگانی، پرسنلی، آموزشی، آماری و . . . پایگاه داده‌ها ایجاد یا خریداری شده است، به طوری که این پایگاه داده‌ها برای مدیران، برنامه ریزان، پژوهشگران و . . . جهت تصمیم‌گیری‌های راهبردی، تهیه گزارش‌های مختلف، توصیف وضعیت جاری خود و . . . می‌تواند مفید باشد. بسیاری از پایگاه داده‌ها چنان گسترش یافته‌اند که شامل چند صد میلیون یا چند صد میلیارد رکورد ثبت شده هستند و امکان تحلیل و استخراج اطلاعات با روش‌های معمول آماری از دل انبوه داده‌ها مستلزم چند روز کار با رایانه‌های موجود است. از این رو به دلیل کاربردی‌تر بودن تکنیک‌های داده کاوی، در سال‌های اخیر این تکنیک‌ها به طور وسیعی در بسیاری از زمینه‌ها بخصوص در مسائل بازرگانی بکار گرفته شده است.

¹ Data Mining

² Knowledge Discovery

یکی از زمینه های مربوطه، محرک های بازاریابی است. در سه دهه گذشته، فروشندگان به شدت بر روی محرک های بازاریابی تکیه کرده اند تا سودشان را به حداکثر برسانند. تحقیقات بازاریابی نشان داده است که محرک های داخلی فروشگاه ها مانند چگونگی چیدمان و نمایش کالاها، ترتیب قفسه ها، فضای داخل فروشگاه و غیره تأثیر شدیدی بر رفتار خرید مشتریان داشته و می تواند موجب حداکثر شدن فروش و یا فروش های ضربدری گردد [29]. برای مثال مطالعات گذشته نشان داده است که از طریق قرار دادن کالاها در مکان های مناسب به گونه ای که پیدا کردن آن ها برای مشتری آسان گردد و یا مجاور قرار دادن کالاهایی، که معمولاً با یکدیگر خریداری می گردند، می توان میزان فروش را حتی تا دو برابر افزایش داد [94]. یکی دیگر از مهمترین مواردی که بر رفتار خرید مصرف کنندگان تأثیر بسزایی دارد، تخفیف کالا است. در واقع تخفیف یک کالا، گاهی اوقات می تواند مهمترین معیار برای خرید یا عدم خرید آن کالا باشد. در مطالعات گذشته [26] در زمینه تخصیص کالاها به قفسه ها³ و در واقع مجاورت کالاها در یک فروشگاه، قیمت و یا تخفیف کالاها لحاظ نگردیده است. از این رو نیاز جهت بررسی رفتار خرید مشتریان با تغییر در چیدمان کالاها براساس تخفیف مشاهده می شود. مدیران فروش باید بتوانند استراتژی هایی را پیاده کنند که نیاز مشتریان را ارضاء کرده و حتی بر روی رفتار خرید آن ها اثر بگذارد. استراتژی های مناسب، اثر قابل توجهی بر روی رفتار خرید مشتریان داشته و می تواند باعث افزایش فروش از طریق ماکزیمم کردن خرید تشویقی و خریدهای ضربدری گردد. در واقع با انتخاب کالاهایی که سودآوری بیشتری دارند (البته انتخاب یک سری از کالاها که ماهیت فروشگاه را نشان می دهند ضروری است) با توجه به درصد تخفیفی که برای آن ها در نظر گرفته شده است و چیدن آن ها در فروشگاه به گونه ای که مشتریان را برای خرید آن ها و خریدهای ضربدری برانگیزد، می تواند حتی موجب افزایش فروش به میزان قابل توجهی شود.

2-1 هدف از اجرای پایان نامه و ضرورت تحقیق

³ Shelf-allocation

تحقیقات انجام شده در این پایان نامه در سه فاز صورت گرفته است. فاز اول ارائه الگوریتم جدیدی برای کاوش قوانین وابستگی چند سطحی در پایگاه های داده تراکنشی با در نظر گرفتن تخفیف کالاهاست، فاز دوم استفاده از نتایج این الگوریتم برای انتخاب کالاها با تخفیفات مناسب می باشد و فاز آخر نیز ارائه الگویی برای چیدمان آن ها با توجه به وابستگی میان آن ها است.

الگوریتم های استاندارد قوانین وابستگی [3, 58] که در تحلیل سبد بازار به کار می روند، عمدتاً در حالتی استاتیک تعیین می کنند که مشتریان چه کالاهایی را با یکدیگر خریداری کرده اند. در واقع، تغییرات دینامیکی نرخ تخفیف روی کالاها و اثرات آن روی فروش نادیده گرفته شده است. مطالعات گذشته [25] که به کشف الگوهایی با اطلاعات تخفیف پرداخته اند، دارای نواقصی می باشند. اولاً، در این تحقیقات فرض شده است که هر کالا در دوره های زمانی مختلف تخفیف های متفاوتی دارد اما این دوره زمانی برای تمامی کالاها ثابت است. در واقع، اگر دوره از قبل تعیین شده ما دو هفته باشد، نرخ تخفیف کالای A و B در این دو هفته تغییر نمی کند. اما در عمل، دوره ای که کالای A تخفیف ثابتی دارد می تواند با کالای B متفاوت باشد. برای مثال در اولین هفته، کالای A می تواند با 10٪ تخفیف فروخته شود و در هفته بعد با 20٪ تخفیف. از سوی دیگر کالای B می تواند در طول این دو هفته با 30٪ تخفیف به فروش برسد. دوماً، این تحقیقات بر روی کشف اثرات تخفیف روی فروش در سطحی منفرد تمرکز کرده اند. به منظور کاربردی تر شدن رویکرد، علاوه بر کشف الگوهایی در سطوح بالایی که به طور کلی ارتباطات را بیان می کنند، کشف قوانین وابستگی در سطوح پایین تر که بسیار خاص تر و دقیق تر است نیز می تواند جالب باشد. همچنین یافتن وابستگی بین کالاها با تخفیف های مختلف در سطوح مختلف در سیاستگذاری بازار مفید است. برای مثال، اگر ماکزیمم تخفیف دسته شیر 30٪ و تخفیف نان سبوس دار سحر 20٪ باشد، 80٪ از مشتریانی که یکی از آیتم های دسته شیر را خریداری می نمایند نان سبوس دار سحر را نیز می خرند. سوماً، در رویکرد گذشته [25] تنها تمرکز روی کشف اثرات مثبت و منفی الگوها بر روی فروش بوده است و قوانین وابستگی بین آیتم ها استخراج نشده اند. در این تحقیق، الگوریتم کامل تری ارائه شده است که نواقص مطالعات قبلی را برطرف می سازد.