

یک سامانه پشتیبانی تصمیم‌گیری بر پایه فن آوری داده‌کاوی،

منطق فازی و الگوی سری‌های زمانی آماری

آرمان دیدنده، دانشجوی مقطع کارشناسی رشته علوم کامپیوتر

دانشگاه صنعتی امیرکبیر تهران

Arman.Didandeh@gmail.com

چکیده

تلاش برای طراحی یک سامانه پشتیبانی تصمیم‌گیری است که قابلیت تصمیم‌گیری در محیط‌های پیچیده را داشته باشد. احتمال عدم قطعیت باعث راهنمایی سامانه مربوطه برای تعیین خودکار جهت برای اتخاذ تصمیمات می‌گردد. سامانه به دانش انتزاعی این اجازه را می‌دهد که در موقعیت‌های متفاوت پردازش، طرح‌ریزی و استنتاج دستورالعمل‌ها، بارها و بارها مورد استفاده قرار گیرد. داده اولیه از منابع مختلف جمع‌آوری شده و پس از پالایش، پردازش به داده‌های منطقی قراردادی برای فرآیند استدلال انجام می‌گیرد. حال این داده‌ها به روش‌های آماری کاربردی نیز پردازش می‌گردند. این سامانه هم‌چنین برای اندازه‌گیری میزان تقاضای بازار در تهیه و تولید یک محصول و یا پشتیبانی از یک مجرای توزیع شده کاربرد دارد.

کلمات کلیدی

داده‌کاوی (Data Mining)، سامانه پشتیبانی تصمیم‌گیری (DSS)،

معرفی

داده‌کاوی فرآیند یافتن روابط، الگوها و روندهای جدید به‌روشنی در مقادیر بسیار داده ذخیره شده در انبارهای داده، با به‌کارگیری هوش مصنوعی و تکنیک‌های ریاضی و آماری است. هم‌چنین داده‌کاوی را می‌توان فرآیند استخراج دانش نهان و ضمنی در حجم بیشماری از داده خام دانست. جایگزینی مناسب برای نام داده‌کاوی، کشف دانش در بانک-های اطلاعاتی (KDD)، کشف دانش و یا تحلیل داده / الگو خواهد بود. حال اهمیت گردآوری داده‌هایی که منعکس-کننده فعالیت‌های تجاری یا علمی برای دستیابی به مزیت‌های رقابتی، کاملاً مشخص است. سامانه‌های بسیار قدرتمندی در سازمان‌های بزرگ و متوسط برای گردآوری داده و مدیریت آن در بانک‌های اطلاعاتی بزرگ و پیچیده به‌کار گرفته می‌شوند. به هر حال تنگنای تبدیل این داده به موفقیت مورد نظر ما، سختی استخراج دانش درباره سامانه‌ای مورد مطالعه از داده مورد نظر است. سامانه‌های پشتیبانی تصمیم‌گیری، ابزارهای رایانه‌ای طراحی شده برای کمک به تصمیم-گیرندگان در راه تصمیم‌گیری‌شان می‌باشند.

سامانه‌های پشتیبانی تصمیم‌گیری به‌طور جدی به مفاهیم عقلانی پیوند خورده‌اند؛ بدین معنی که تمایل به انجام کارها به‌روش مستدل و منطقی برای ایراد تصمیم مناسب وجود دارد. برای ایراد تصمیمات اساسی و کلیدی در یک سازمان درگیر با مسائل محصولات و خدمات، موارد زیر بسیار مهم و اساسی می‌نمایند:

- توزیع محصولات با به‌کارگیری مجاری مختلف توزیع
- محاسبات برون‌داد (خروجی) بر مبنای زمان و فضای متفاوت
- پیش‌بینی روند برون‌داد محصولات یا خدمات به‌صورت منحصر به‌فرد و در چهارچوب زمانی برآوردی (تخمینی)

- برنامه زمان‌بندی محصول بر پایه تقاضا، ظرفیت و منابع
- هدف و رسالت اصلی کار، طراحی سامانه‌ای بر اساس تصمیمات پویا است که به خصوصیات انحصاری دوره زندگی (Lifecycle) محصول وابسته باشد. تحلیلات گرافی بسیاری برای دادن روند فکری تسهیلی و پیش‌بردی به تحلیل الگوی محصول یا خدمت انجام شده است. سامانه از جنبه‌های محلی و جهانی برای چندین بار مورد بازدید و مرور قرار گرفته است.

توانایی داده‌کاوی

به‌طور کلی می‌توان از فن‌آوری داده‌کاوی در شاخه‌های زیر سود جست:

- تجارت
- مدیریت بازار
- تحلیل بازار
- پیش‌بینی آب و هوا
- ارزیابی‌های محیطی
- و سامانه‌های زیست‌شناختی (بیولوژیکی)

داده‌کاوی: فرآیند کشف دانش از داده (KDD)

فرآیند اکتشاف دانش از داده‌ها نیازمند دانش پیشین و مناسب در ارتباط با برنامه‌های کاربردی و اهداف آن‌ها است؛ ایجاد یک مجموعه داده‌ای بزرگ، پیش‌پردازش داده، پالایش و تصفیه و شفاف‌سازی، تبدیل داده، شناسایی ویژگی‌های ابعادی کارا و سودمند. هم‌چنین این فرآیند دربردارنده طبقه‌بندی، اجتماع، پس‌گشت (رگرسیون)، خوشه‌بندی و خلاصه‌سازی می‌باشد. مهم‌ترین عامل این فرآیند، انتخاب الگوریتم داده‌کاوی است. مرحله آخر شامل ارزیابی الگو است که به معنی تجسم فکری، تبدیل و حذف الگوهای زاید با به‌کارگیری دانش استخراجی از فرآیند می‌باشد.

فن‌آوری داده‌کاوی و سامانه: متدهای داده‌کاوی شامل شبکه‌های عصبی، برنامه‌نویسی تکاملی، برنامه‌نویسی بر مبنای ساختار حافظه، ساختار درختی تصمیم‌گیری، الگوریتم‌های پیدایشی (Genetic Algorithms) و متدهای غیر خطی رگرسیون است. هم‌چنین در این کار منطق فازی - که یک سری قوانین فوق مجموعه‌ای برای منطق بولی (Boolean Logic) با توسعه مفاهیم درستی (Truth) و نادرستی (Falsity) جزئی (Partial) و قرار دادن آن‌ها بین درستی کامل و نادرستی کامل است - بسیار کاربرد دارد.

روال کار

شروع کار با شناسایی نیازهای کاربر است. ابهامات این موضوع در زیر آمده‌اند:

- انواع داده‌ای در انبار داده (Data Warehouse).
 - کمبودهای ناشی از تخصیص دوره‌ای منابع.
 - ویژگی مقیاس‌ها، معیارها و ضوابطی که به تصمیم‌گیری کمک می‌کنند.
- ردیابی کارآیی پایین در مراحل مختلف، باعث پیگیری مداوم اعمال مشخصه تکنیک برای ایراد تصمیمات است. برای تحلیل نیازها، نیازهای کاربر به نیازهای قابل آزمودن بدل می‌شود. عوامل پایه بسیاری برای اندازه‌گیری پیش‌بینی بهینه و ایده‌آل وجود دارند، که درمورد بحث ما می‌توان به موارد زیر اشاره نمود:

- کارآیی نمایش گاهی (Showroom Performance)
 - کارآیی محصول (Product Performance)
 - کارآیی گروه محصولی (Product Group Performance)
 - دوره زندگی محصول (Product Lifecycle)
 - محیط (Environment)
 - دستورالعمل های کاربری (User Instruction)
 - اجبارات و تعهدات لازم (Obligation)
- در زیر می توانید شبهه کد مورد نظر را مشاهده نمایید:

```

if(findHistory(single product history) is TRUE) then
    pMeasure(single product);
else if(findHistory(product group history) is TRUE) then
    pMeasure(product group);
else
    pMeasure(showroom history);

```

این شبهه کد در زیر به جدول دارای ساختار درختی تصمیم گیری بدل شده است. علامت X در جدول به معنای عدم دخیل بودن عامل در تصمیمات است.

Table Input: Unit Product		Table Output: Nearest Group of Showroom		
Search for Nearest Product Group				
History	Single Product	Product Group	Showroom	Action
Rule 1	F	F	F	Showroom
Rule 2	F	F	T	Showroom
Rule 3	F	T	X	Group
Rule 4	T	X	X	Single Product

محاسبات بهترین کارآیی: عوامل پایه برای محاسبه ضوابط کارآیی پایه یافت شده اند. عوامل اصلی کارآیی پایه عبارتند از:

- کارآیی عامل فروش
- کارآیی عامل موجودی انباشته
- تعداد فروش طراحی ها
- کارآیی فروش طراحی های جدید

کاربرد منطق فازی: با به‌کارگیری ملاک‌های فازی در بررسی و محاسبات کارآیی می‌توان به قطعیت دادن بیش‌تر به تصمیمات اتخاذ شده کمک شایانی نمود. استفاده از منطق فازی باعث تغییر میزان محاسبه‌ای تمام عوامل کارآیی به مقداری بین صفر و یک خواهد شد. تکنیک مورد بحث در زیر ارایه شده است:

فازی کردن عامل فروش: میزان فروش به عنوان A در نظر گرفته می‌شود. کم‌ترین مقدار را X و بیش‌ترین مقدار را Y در نظر می‌گیریم.

	Rule#1	Rule#2	Rule#3
Gross Sale	< X	[X : Y]	> Y
Gross Sale Factor (A)	0	(Sale - X) / (Y-X)	1

فازی کردن عامل انباشته اضطراری: تعداد انباشته موجود را B در نظر می‌گیریم. کم‌ترین مقدار را X و بیش‌ترین مقدار را Y در نظر می‌گیریم.

	Rule#1	Rule#2	Rule#3
Stock Lot	> X	[X : Y]	< Y
Stock Lot Factor (B)	0	(X - Stock Lot) / (X - Y)	1

فازی کردن عامل تعداد مدل‌های طراحی شده: تعداد فروش طراحی‌ها را C در نظر می‌گیریم. کم‌ترین مقدار را X و بیش‌ترین مقدار (که برابر ۱۵۰۰ است) را Y در نظر می‌گیریم.

	Rule#1	Rule#2	Rule#3
No. Of Design Sale	< X	[X : Y]	> Y
No. Of Design Sale Factor (C)	0	(No. of Design Sale - X) / (Y-X)	1

فازی کردن عامل ترویج و تبلیغ محصول جدید: درصد فروش طراحی‌های جدید را D در نظر می‌گیریم. کم‌ترین مقدار (که برابر ۱۰ درصد است) را X و بیش‌ترین مقدار (که برابر ۸۰ درصد است) را Y در نظر می‌گیریم.

	Rule#1	Rule#2	Rule#3
% of New Design Sale	< X	[X : Y]	> Y
% of New Design Sale Factor (D)	0	(% of New Design Sale - X) / (Y-X)	1

تابعی برای ارزیابی کارآیی ناخالص: این تابع با اجرای هر یک از جداول فوق، مقداری بین صفر و یک به دست می‌آورد. اگر حتی یکی از مقادیر برابر صفر بود، محاسبات توزیع متوقف خواهد شد:

function fuzzyPerformanceForShowRoom(ShowRoom i)

var

percentFit, iterator: integer;
sr: array[1..n] of integer;
factorResult: array[1..4] of real;
flag: Boolean;

Begin

flag = true;
factorResult = executeAllFactors();
for(iterator : 1 to 4)
 if(factorResult[i] < 0) then
 flag = false;
if(flag) then
 percentFit = ((A + B + C + D) / 4) * 100;
else
 percentFit = 0;
return percentFit;

End;

محاسبه کارآیی PLC: بر مبنای عامل فازی در بازه‌های زمانی مختلف، محاسبات PLC با استفاده از معادلات زیر انجام می‌شود:

$$S_t = \alpha \frac{y_t}{I_{t-L}} + (1 - \alpha)(S_{t-1} + b_{t-1}) \text{ Overall Smoothing} \quad (1)$$

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \lambda)b_{t-1} \text{ Trend Smoothing} \quad (2)$$

$$I_t = \beta \frac{y_t}{S_t} + (1 - \beta)I_{t-L} \text{ Seasonal Smoothing} \quad (3)$$

$$F_{t+m} = (S_t + mb_t)I_{t-L+m} \text{ Forecast} \quad (4)$$

که در آن y میزان مشاهده، S میزان مشاهده بدون ایراد، b عامل روند کار، I شاخص فصلی، F پیش‌بینی در m بازه آینده، t شاخص معین‌کننده یک بازه زمانی و α ، β و γ ثابت‌هایی هستند و به‌گونه‌ای انتخاب خواهند شد که ماکزیمم خطای ما مینیمال گردد.

بحث

در این پروژه، داده‌ها از انبار داده استخراج می‌گردند. سپس داده استخراجی پالایش گشته و برای فرآیند استدلال به داده رسمی منطقی و عقلانی پردازش می‌شود. به‌عنوان مرحله مقدماتی فرآیند استدلال، منطق فازی بر روی تمام مشخصات و حقایق رفتاری مستقل هر مجموعه منحصر به فرد داده اعمال می‌گردد، که برای تحلیل کارآیی نمایشگاهی برای محصولات منحصر به فرد یا ضوابط محاسباتی محصولات گروهی عبارتند از: فروش ناخالص، انباشته اضطراری، میزان فروش طراحی‌ها، فروش محصول جدید و مانند آن‌ها. سه لایه پنهانی در منطق فازی موجود است که بر ساختار

درختی تصمیم‌گیری پویای آن اثرگذار می‌باشد. پس از اجرای منطق فازی، تمام ضوابط محاسبه کارایی در فضای دامنه-ای مشخص، مثلاً بین صفر و یک قرار خواهند گرفت. سپس ضوابط کارایی تحت متد میانگین‌سازی و با یک قانون ساده آمیخته می‌شوند تا اگر تمام ضوابط بزرگتر از صفر بودند، میانگین آن‌ها در نظر گرفته شود و در غیر این صورت مقدار صفر برای ضابطه کلی تثبیت گردد. نتیجه کارایی به‌وسیله فرمول نمای هموار سه‌گانه مدل سری‌های زمانی پیش‌بینی می‌گردد که در آن داده‌های اخیر بسیار اثرگذارتر از داده‌های به نسبت قدیمی‌تر باشند. این نتیجه مداوماً با محیط در حال تبادل شرایط است تا نتیجه نهایی را برای تعیین محصول و تخصیص منابع بر اساس الگوی کارایی بیابد.

راه‌های بسیاری برای این پیش‌بینی وجود دارد. متدهای پیش‌بینی سری‌های زمانی بر مبنای تحلیل داده‌های تاریخی می‌باشند. این محاسبات بر اساس مشاهدات در بازه‌های مختلف متواتر انجام می‌گیرد. این پنداشت را به‌وجود می‌آورد که الگوهای پیشین موجود در داده‌ها می‌تواند برای تخمین آینده به‌کار رود. راه‌های عمده و معروف و محبوب پیش‌بینی عبارتند از: میانگین‌های ساده و وزن‌دار در فرمول‌های نمای هموار سه‌گانه، دوگانه و یگانه. در بین این‌ها، فرمول سه‌گانه بسیار حساس‌تر و موثرتر است زیرا دارای سه عبارت کنترلی است: مینا، روند و تأثیرات فصلی. مقدار ثوابت می‌تواند باعث مدیریت و کنترل و یا حتی تحریک اثر تقسیمات زمانی در یک دوره مشخص زمانی، مثلاً بازه‌های فصلی یا ماهانه در مدت چندین سال گردد. حساس‌ترین و بحرانی‌ترین قسمت کار عواملی است که ممکن است برای هر کارایی نمایشگاهی یک شناسه منحصر به‌فرد باشند. تمام سامانه با مدل هسته‌ای مهندسی نرم‌افزاری که در این‌جا به آن اشاره می‌کنیم طراحی می‌شود.

فرآیند با بازنگری ابهامات آغاز شده، پس از تحلیل و نگرشی به اقلام حیاتی نیازمندی‌ها، کار تحلیل تحت فاز شناخت نیازها برای تصویر کردن یک مدل کاری برای سامانه تحلیلی جزئی انجام می‌گیرد. مطالعات امکان‌سنجی تحت این مدل سامانه‌ای انجام می‌شود. برای دستیابی به نتایج دقیق‌تر، توده‌های داده‌ای متنوعی برای احراز حقایق مسلم در مورد عملیات تحلیلی هموار مورد نیاز است.

استنتاج

یک سامانه دانش‌محور برای تصمیم‌گیری تحت شرایط عدم قطعیت حداکثری ارایه شد. مشخصه برجسته این سامانه دربردارندگی توزیع بهینه، پایه اطلاعاتی روزانه سناریوی کارایی، الگوهای توزیع و بازتوزیع، تحلیل کارایی نمایشگاهی محصول و قضاوت روی بازار است. سامانه بر اساس مفاهیم سامانه باز و بر مبنای نسبت مشتریان بازار، چرخه حیات محصول، محیط بازار نیچ و پراکندگی و امکان بی‌شمار انتخاب ساخته می‌شود.

منابع

1. Decision Support Systems:
<http://en.wikipedia.org/wiki/DSS>
2. Md. Shamim Ahsun & Dr. Shahida Rafique, Data warehousing & data mining; Prediction DSS for renewal energy in Bangladesh. In proceeding of the conference on "Energy for Sustainable Development: Technology Advances & Environmental Issues" Cairo, Egypt, December 2004.
3. Mark A. Hall, Geoffrey Holmes, Benchmarking Attribute Selection Techniques for Discrete Class Data Mining, IEEE Transactions on Knowledge and data engineering Vol. 15, No. 3, May/June 2003.

4. Dr. Shahida Rafique, Khan R, Md. Jabiullah Ismail, Begum Q N "Time-series Analysis of Solar Radiation in Bangladesh" in the proceedings on the 2nd Int. Seminar on Renewable Energy for Poverty Alleviation, Dhaka, Bangladesh. Pages: 116 to 122.
5. Auroop R Ganguly, Amar Gupta, Shiraj Khan, Data Mining & Decision Support for Business & Science, Encyclopedia of data mining & warehousing, Idea Group Inc., June 2005
6. Nada Lavarac, Data Mining & Decision Support: A note on the issues of their integration & their relation to expert systems, J Stefan Institute, Slovenia