



ارائه شده توسط:

سایت ترجمه فا

مرجع جدیدترین مقالات ترجمه شده

از نشریات معتبر

کاوش کارآمد قوانین روابط فازی از جریان های فراگیر داده ها

چکیده

به دلیل توسعه در فناوری، تعداد برنامه هایی نظیر گوشی های هوشمند موبایل، شبکه های حسگر و ابزار های جی پی اس تولید حجم عظیمی از داده های فراگیر در شکل جریان یا استریم می کنند. جدا از داده های موجود در پایگاه های اطلاعاتی متعارف استاتیک، جریان های فراگیر داده ها به طور پیوسته با سرعت بالا با مقدار زیاد حاصل می شوند و توزیع داده ها نیز روز به روز تغییر می کند. رسیدگی به اطلاعات مفید و استخراج آن ها از این داده ها می تواند یک چالش واقعی باشد. این خود می تواند منجر به بروز مسائل جدیدی شود که بایستی هنگام توسعه فنون کاوش قواعد ارتباطی برای این داده ها در نظر گرفته شود. لازم به ذکر است که در دنیای واقعی، این داده ها به شکل دو دویی و یا عددی نشان داده نمی شوند بلکه بایستی به شکل کمی بیان شوند. از این روی، استفاده از مجموعه های فازی برای نشان دادن این مقادیر می تواند مفید و مناسب باشد.

در این مقاله، مسئله کاوش قواعد ارتباط فازی از جریان داده های فراگیر مطالعه می شود و روش جدید FFP_USTREAM توسعه داده می شود. این روش، مفاهیم فازی را با جریان داده های فراگیر ترکیب کرده و از روش پنجره کشویی برای کاوش قواعد ارتباط فازی استفاده می کند. به علاوه، پیچیدگی و کارایی این روش بحث می شود. نمونه هایی از مجموعه داده های واقعی برای تست روش های پیشنهادی استفاده می شود. به علاوه مسائل مربوط به تحقیقات نیز پیشنهاد می شوند.

لغات کلیدی: داده کاوی، قواعد ارتباط فازی، مجموعه های فازی، جریان داده ها، جریان های فراگیر

1-مقدمه

برنامه های جدید اخیر نظیر پایش ترافیک شبکه، تحلیل داده های شبکه حسگر، کاوش جریان کلیک جهانی، اندازه گیری مصرف برق و مکان یابی دینامیکی نوسانات بازار سهام، لزوم مطالعه انواع جدید داده ها را تاکید می کند. این موسوم به داده های جریان است که به صورت جریان نامحدود و پیوسته از اطلاعات بر خلاف جریان های داده های ذخیره شده آماری محدود است. داده کاوی جریان، فرایند استخراج ساختار های دانش از داده های پیوسته می باشد.

انتشار سیستم های جریان داده، شبکه های بی سیم و ابزار های سیار موجب می شود تا لزوم استفاده از ابزار های تحلیل کارآمد داده ها در مورد کسب اطلاعات از جریان های داده پیوسته در نظر گرفته شود. کاوش جریان داده های فراگیر، فرایند کشف الگو بر روی ابزار های فراگیر و سیار است. این خود نشان دهنده نسل بعدی از سیستم های داده کاوی است که از نیاز عای اطلاعاتی زمانی و مکانی کاربران موبایل پشتیبانی کرده و موجب تسهیل داده کاوی هر زمان هر جا می شود.

منطق فازی، نوعی منطق مورد استفاده در هوش مصنوعی است. این اشاره به منطق چند مقدار دارد. به جای داشتن دو مقدار (حقیقی و کاذب)، طیف پیوسته ای از مقادیر حقیقی محتمل (2) وجود دارند. در منطق فازی، هر فرض یک گزاره ای است که به عدد بین 0 (کاذب) و 1 (حقیقی) نسبت داده می شود نظیر گزاره ای که موسوم به فرض فازی است. منطق فازی، ابزاری قوی برای طبقه بندی مفهوم به شکل انتزاعی با معرفی ابهام است.

بسیاری از برنامه های جریان داده وجود دارند که نیاز به کاوش قواعد ارتباطی دارند نظیر پایش ترافیک شبکه و تحلیل جریان کلیک وب. هدف این برنامه ها کشف روابط مهم میان ایتم ها است زیرا حضور برخی از ایتم ها نشان دهنده حضور برخی دیگر است.

رویکرد قواعد روابط فازی می تواند نتایج داده کاوی را با تخصص انسانی و دانش قبلی به شکل قواعد ترکیب کند تا به دسته ای از طبقه بندی های جریان داده ها برسد. دیگر مزیت رویکرد منطق فازی این است که نتایج طبقه بندی را می دهد که شامل درجه ای از احتمال است.

این مقاله، کارایی و اثر بخشی کاوش قواعد ارتباط فازی را از جریان داده های فراگیر اثبات می کند. این خود در بخش های بعدی بررسی می شود. برای این منظور، بخش باقی مانده مقاله به صورت زیر سازمان دهی شده است: کار مربوط به کاوش قواعد ارتباطی فازی و کاوش جریان داده های فراگیر مرور و در بخش 2 خلاصه شده است. ک روش کاوش قواعد ارتباط فازی از داده های فراگیر در بخش 3 پیشنهاد شده و پیچیدگی آن در بخش 4 تحلیل شده است. به علاوه، نتایج آزمایشی در بخش 5 بررسی شده است. مسائل مربوط به تحقیقات آینده در بخش 6 ارائه می شود.

2- کار های مربوطه

این مقاله مربوط به زمینه های تحقیقات بین رشته ای مختلف است. دو موضوع اصلی این کار ارایه می شود: روش های کاوش جریان داده های فراگیر برای تحلیل موثر داده های پیوسته و الگوریتم های کاوش قواعد ارتباط فازی. این دو زمینه نیز بررسی می شوند.

2-1 کاوش جریان داده های فراگیر

روش بر اساس مجموعه داده های ذخیره شده آماری، در بسیاری از زمینه ها کاربرد زیادی ندارد. این موارد شامل تحلیل شبکه بی سیم، تشخیص نفوذ، تحلیل بازار بورس، تحلیل داده های شبکه حسگر می باشد و به طر کلی در شرایطی که در آن هر اطلاعات قابل دسترس بایستی برای تشخیص نفوذ استفاده شود. این شرایط مستلزم استفاده از الگوریتم های جدید می باشد که قادر به حل بسیاری از مشکلات مربوط به داده ها می باشند که در جدول 1 نشان داده شده است.

| | |
|--------------------------|--|
| <i>Data localization</i> | |
| Centralized | A single entity can access every data |
| Distributed | Each node can access just a part of the data |
| Homogeneous | ... data related to the same entity (e.g. people) are owned by just one node |
| Heterogeneous | ... data related to the same entity (e.g. people) may be spread among several nodes |
| <i>Data evolution</i> | |
| Static | Data are definitively stored and invariable (e.g. related to some past and concluded event) |
| Incremental | New data are inserted and access to past data is possible (e.g. related to an ongoing event) |
| Evolving | The dataset is modified with either updates, insertions or deletions, and access to past data is possible |
| Streaming | Data arrives continuously and for an indefinite time. Access to past data is restricted to a limited part of them or summaries |

جدول 1

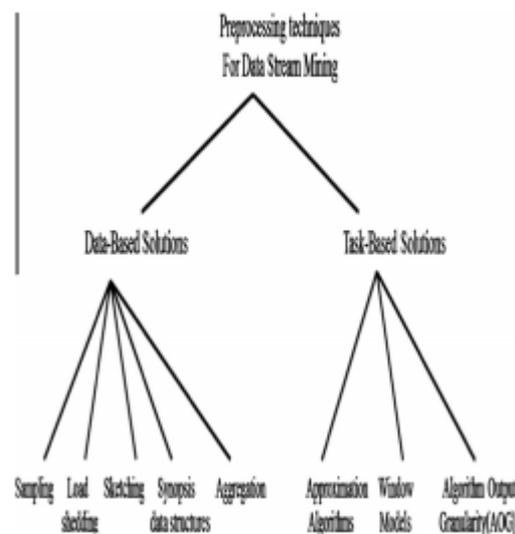
داده کاوی فراگیر یک فرایند بحرانی زمانی کشف الگو در جریان داده ها در محیط بی سیم است (4). استفاده گسترده از دستگاه های موبایل، با افزایش قابلیت محاسباتی، منجر به بروز پارادایم های محاسباتی فراگیر شده است. این پارادایم ها موجب تسهیل دسترسی به داده ها و اطلاعات از طریق کاربران موبایل با ابزار های دستی شده است (5). UDM فرایند تحلیل داده ها از منابع ناهمگن و توزیعی با دستگاه های موبایل یا درون شبکه های حسگر است که در آن داده ها به طور پیوسته به درون دستگاه وارد می شوند و دارای محدودیت های زمانی است که مستلزم تحلیل همه زمان همه جا است (6).

در زیر بخش های زیر، پیش پردازش مورد نیاز برای جریان داده ها ارایه می شود.

1-1-2 روش های جریان داده ها

مسائل و چالش های پژوهشی به وجود آمده در جریان های داده کاوی دارای راه حل هایی هستند که از رویکرد های محاسباتی و آماری بهره می برند. این روش ها را می توان طوری طبقه بندی کرد که شامل انواع داده محور و وظیفه محور هستند. این طبقه بندی در شکل 1 (7) نشان داده شده است. در راه حل های داده محور، ایده اصلی بررسی زیر مجموعه ای از مجموعه داده های کل یا تبدیل داده های افقی یا عمودی به نمایش داده های با اندازه کوچک تر است. از سوی دیگر، در راه حل های وظیفه محور، روش های نظریه های محاسباتی برای رسیدن به روش های کارآمد زمانی و مکانی استفاده شده اند.

1-1-1-2 روش های داده محور: روش های داده محور اشاره به خلاصه سازی مجموعه داده های کل یا انتخاب زیر مجموعه ای از جریان ورودی دارد. نمونه برداری و روش های طرح بندی بسیار رایج می باشند (7).



شکل 1- طبقه بندی روش های پیش پردازش جریان داده ها

2-1-2 قوانین ارتباط جریان داده ها

تعدادی از الگوریتم ها برای استخراج دانش از اطلاعات در حال جریان پیشنهاد شده اند. این موارد شامل خوشه بندی، طبقه بندی، شمارش فرکانس و تحلیل سری های زمانی است. از آن جا که تعداد برنامه های مربوط به جریان داده کاوی به شدت رشد می کند، نیاز روز افزونی به انجام کاوش قواعد ارتباطی بر روی داده های جریان

مشاهده می شود. یک کاربرد نمونه از کاوش قواعد ارتباط جریان، برآورد داده های مفقود در شبکه های حسگر می باشد(8). مثال دیگر، پیش بینی برآیند فراوانی و فرکانس جریان بسته اینترنت است(9).

در پروژه MAIDS(10)، این روش برای یافتن برخورد های هشدار دهنده از جریان داده ها استفاده می شود. کاوش قواعد ارتباطی را می توان برای پایش جریان تولید جهت پیش بینی شکست و تولید گزارش های مبتنی بر جریان های لوگ وب استفاده کرد(11).

چندین الگوریتم الگوی متوالی (12-18) در منابع پیشنهاد شده است. بیشتر این روش ها تولید نتایج تقریبی کرده، داده های خاص را کشف کرده و نتایج دقیقی ارائه می کنند. این ها در جدول 2 مقایسه شده اند.

2-2 کاوش قواعد ارتباط فازی

بیشتر مطالعات نشان داده اند که چگونه داده های تراکنش دودویی را می توان بررسی کرد. با این حال، داده های تراکنش در برنامه های واقعی متشکل از مقادیر فازی و کمی است. از این روی، طراحی الگوریتم های داده کاوی پیچیده امکان استفاده از انواع مختلف داده ها را برای کارکنان در این زمینه تحقیق در اختیار می گذارد. این الگوریتم های قواعد ارتباط فازی (20-22) اخیراً در منابع ارائه شده اند این موارد در جدول 3 مقایسه شده اند.

2-3 لزوم توسعه کار های موجود

مطالعه جداول 2 و 3 نشان دهنده لزوم توسعه یک روش جدید برای کاوش قواعد ارتباط فازی از جریان داده های فراگیر می باشد: این روش از خصوصیات زیر استفاده می کند

- توانایی رسیدگی به جریان های پیوسته جریان های داده
- توانایی پیش نویس مفهومی داده ها طی زمان
- توانایی اندازه حافظه محور، مناسب برای کاربرد های فراگیر
- تسهیل تحلیل داده ها و پشتیبان تصمیم گیری سریع برای کاربران با اسکن جریان های داده
- توانایی تعیین مجموعه های فازی و توابع عضویت
- تعیین پشتیبانی مبتنی بر کاربر
- ارائه نتایج صحیح

در بخش های زیر، یک روش جریان فراگیر الگوی فازی کارآمد که مطابق با خصوصیات فوق است

3-روش پیشنهادی: FFP- USTREAM

روش پیشنهادی متشکل از چهار مرحله می باشد که در الگوریتم a-1 نشان داده شده است. این موارد در زیر دیده می شود

3-1 مرحله 1- تعیین پنجره های کشویی

برای رسیدگی به جریان های داده های فراگیر، مدل پنجره کشویی در الگوریتم a-2 برای یافتن الگوهای فازی دقیق استفاده می شود. مرحله نخست، تعیین اندازه پنجره کشویی است. تراکنش های قدیمی زمانی منقضی می شوند که تراکنش های جدید به پنجره فعلی برسد. اندازه پنجره بستگی به برنامه و منابع سیستم (19) دارد. در این کار، اندازه پنجره به طور ثابت در نظر گرفته می شود. اندازه پنجره متغیر برای کار آینده حفظ می شود.

یک نمونه از پنجره کشویی در جریان های داده های فازی در شکل 2 نشان داده شده است. در این مثال، اندازه پنجره و تعداد تراکنش ها به مقدار 2 تنظیم می شود. (A:5) به معنی تعداد واحد های خریداری شده ایتm A است.

3-2 مرحله 2- فازی سازی

فازی سازی مقادیر جریان داده ها شامل دو مرحله است. در ابتدا، مجموعه های فازی برای متغیر ها تنظیم شده و آن ها را با شکل زبانی نشان داده می شود (23). دوما، توابع عضویت برآورد می شود (24). در عمل، توابع عضویت می تواند دارای انواع مختلف چندگانه نظیر شکل موجی مثلثی، دوزنقه ای، گوسی، زنگوله ای، سیگموییدی و منحنی های S باشد.

Table 2 Comparative analysis of frequent pattern algorithms [19].

| Algorithm name | Window model | Algorithm type | Update rate | Merits | Limitations |
|--|--------------------|----------------|------------------|---|---|
| Lossy counting [12] | Landmark | Approximate | Batch wise | No false negatives in results | Setting up of relaxed minimum support threshold leads to dilemma |
| Data Stream Mining Frequent Itemsets (DSM-FI) [13] | Landmark | Approximate | Batch wise | Compact tree structure has been designed to store the frequent patterns | It needs more tree traversals for the frequency count |
| Compact Prefix tree Structure (CPS) [14] | Sliding window | Exact | Batch wise | It maintains the frequency count lists at the last node, which can reduce the size of the prefix tree | Additional computational cost needed for restructuring the tree after every pane insertion. |
| Frequent Pattern (FP-Stream) [15] | Tilted time window | Approximate | Batch wise | It extracts complete set of frequent patterns using time sensitive data streams | FP-Stream tree becomes very large with time |
| Weighted sliding Window (WSW) [16] | Sliding window | Exact | Batch wise | A single pass algorithm was developed to discover the frequent itemsets | Weights of each window affected the mining results. So, user should specify the reasonable weight for each window |
| Weighted Support Frequent Pattern (WSFP) [17] | Sliding window | Exact | Transaction wise | It collects the important recent frequent patterns with limited memory space | Initial setting of normalized minimum and maximum weight is given as random |
| Variable Sliding Window VSW [18] | Sliding window | Exact | Batch wise | Obsolete transactions are deleted with respect to calculated formula | The prefix tree becomes very large where there is no changes occurred in the frequent patterns in processing |

جدول 2

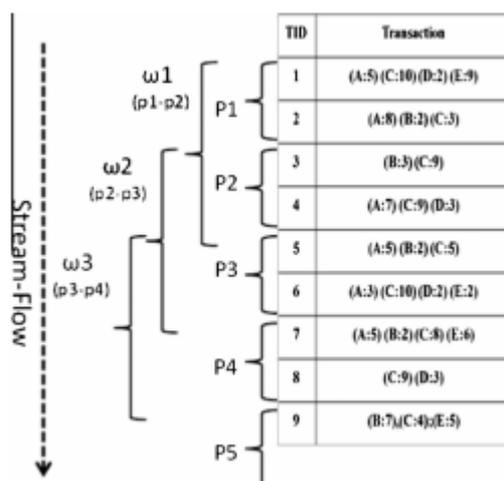
| Algorithm | Mining task | Advantages | Disadvantages |
|---|---|---|---|
| F-APACS [20] | <ul style="list-style-type: none"> - Transformed quantitative attribute values into linguistic terms - Used the adjusted difference analysis to find interesting associations among attributes - Based on the a priori approach | <ul style="list-style-type: none"> - Discover both positive and negative associations - Avoid the use of some user-supplied thresholds | <ul style="list-style-type: none"> - Membership function must to be given - Cause iterative database scans - High computational costs - Does not handle data streams |
| Fuzzy Frequent Pattern Tree (FFPT) [21] | <ul style="list-style-type: none"> - Update the fuzzy sets at all the phases of the algorithm - Preserve the original information of the data sets - Based on Frequent Pattern Tree approach. | <ul style="list-style-type: none"> - Avoid the candidate generation phase - Avoid the repetitive scanning of the original database | <ul style="list-style-type: none"> - Could not generate the membership value of an itemset - Local frequent fuzzy regions were used to construct the FFPT - Tree structure was loose and huge - Does not handle data streams |
| Compressed Fuzzy Frequent Pattern Tree (CFFPT) [22] | <ul style="list-style-type: none"> - Integrates the fuzzy-set concepts and the FP-tree-like approach to efficiently find the fuzzy frequent itemsets from the quantitative transactions - Create CFFP-tree - CFFP-growth mining algorithm is proposed to derive fuzzy frequent itemsets from the constructed CFFP tree | <ul style="list-style-type: none"> - The node number in the tree can be reduced - The CFFP tree Structure is tighter and more compressed - Can process quantitative transactions | <ul style="list-style-type: none"> - Additional array stored in each node - The proposed approach will spend more execution time because the former will need to transfer the quantitative values into fuzzy sets - The database is assumed static - Minimum support threshold is given - Does not handle data streams |

جدول 3

3-3: مرحله 3- ساخت ساختار درختی

به منظور استنباط قوانین ارتباط فازی از جریان های داده های فراگیر، درخت الگوی فراوانی فازی دینامیک ساخته می شود. ساختار پایه درخت DFFP مشابه با درخت الگوی متوالی است. با این حال، مقدار عضویت به هر گروه افزوده می شود که در الگوریتم A-3 نشان داده شده است.

به منظور این که روش حاضر برای ابزار های فراگیر با منابع محدود مناسب باشد، یک الگوریتم a-4 برای حذف الگوهای قدیمی افزوده می شود و درختان باز سازی شده تغییر می یابند.

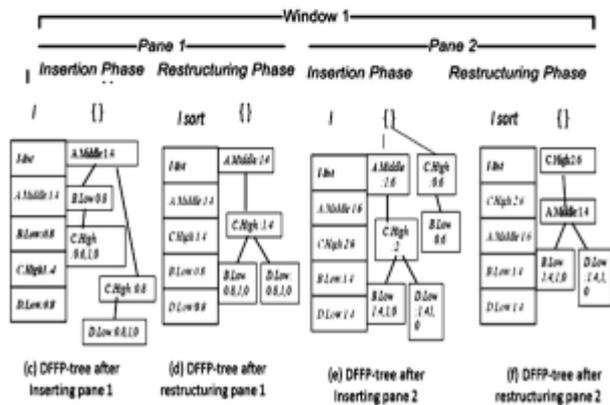
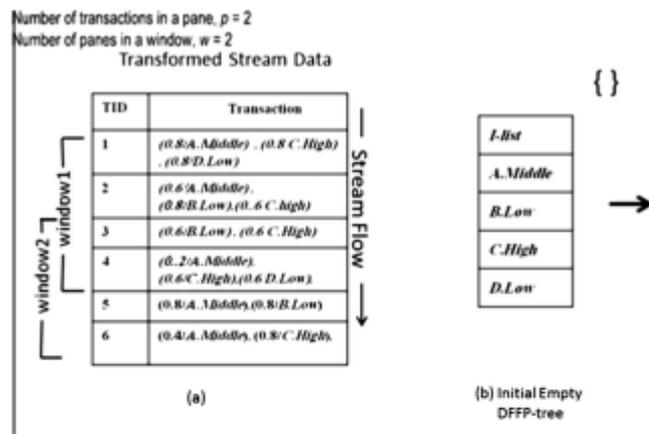


شکل 2- جریان داده ای مبتنی بر پنجره کشویی

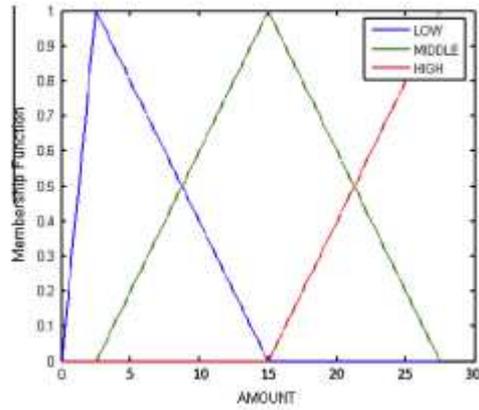
روش ساخت گام به گام در شکل 3 نشان داده شده است. به علاوه تحلیل های قیاسی بین ساختار های درختی DFFP در جدول 4 نشان داده شده است. 3-4: مرحله 4- استخراج قوانین ارتباط فازی

3-4 مرحله 4- استخراج قوانین ارتباط فازی

وقتی که درخت DFFP ساخته شد، الگوریتم a-5 برای تولید مجموعه کاملی از الگوهای فازی از پنجره استفاده می شود. از این روی، قوانین ارتباط فازی بر اساس بازه اطمینان خاص با استفاده از الگوریتم A-6 استنباط می شود.



شکل 3- ساخت درخت الگوی فازی دینامیک



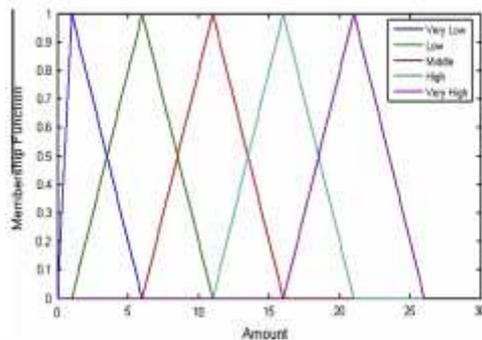
شکل 4- توابع عضویت با سه مجموعه فازی منطقه ای

در این بخش، پیچیدگی FFP-USTREAM تجزیه تحلیل می شود. اصطلاحات زیر برای این بحث تعریف می شوند

P: تعداد تراکنش ها در هر پنجره

S: تعداد مجموعه های فازی

A: تعداد صفات مجزای هر مطالعه



شکل 5- توابع عضویت با مجموعه های فازی مناطق

Table 5 Dataset characteristics.

| Data set name | No. of trans. | No. of items | Total length | Average length % |
|----------------|---------------|--------------|--------------|------------------|
| Accidents [26] | 30571 | 342 | 3164207 | 30.2% |
| Retail [27] | 29387 | 13737 | 1362940 | 0.34% |

جدول 5

این بحث به دو بخش تقسیم می شود: پیچیدگی زمان اجرا و پیچیدگی حافظه

1-4 پیچیدگی زمان اجرا

می توان نشان داد که این پیچیدگی به سه مرحله اصلی تقسیم می شود:

- فرایند فازی سازی که نقشه صفات مقادیر کمی را با مجموعه های فازی تهیه می کند. تعداد نقشه ها به ترتیب $O(|S| * |P| * |I|)$ است.
- ساخت درخت DFFP: این خود پایه و اساس روش پیشنهادی است. پیچیدگی به ترتیب $O(|P| * |I|)$ می باشد که در بر گیرنده مقادیر گره و یا ایجاد یک گره جدید است.
- استخراج قواعد ارتباط فازی، پیچیدگی آن به ترتیب $O(|P| * |I|)$ نشان دهنده تعداد مقایسات است.

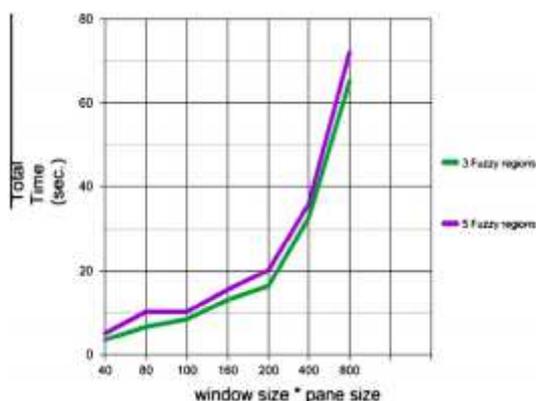
Table 6 Runtime distribution with variation of fuzzy regions.

| Data sets | Window size | Pane size | | | | | |
|-------------------|-------------|---------------------|--------------------|---------------------|--------------------|---------------------|--------------------|
| | | 20 | | 50 | | 100 | |
| | | Runtime in sec. | | Runtime in sec. | | Runtime in sec. | |
| | | Three fuzzy regions | Five fuzzy regions | Three fuzzy regions | Five fuzzy regions | Three fuzzy regions | Five fuzzy regions |
| Accidents (Dense) | 2 | 3.4788 | 5.1387 | 8.3928 | 10.23 | 16.4872 | 19.152 |
| | 4 | 6.552 | 10.24 | 16.38 | 20.34 | 32.532 | 35.858 |
| | 8 | 13.1678 | 15.542 | 32.5574 | 35.94 | 65.0862 | 71.988 |
| Retail (Sparse) | 2 | 0.421 | 1.1644 | 1.53252 | 2.3394 | 4.09 | 5.14 |
| | 4 | 1.404 | 2.3213 | 2.5741 | 4.8373 | 8.2 | 9.677 |
| | 8 | 3.0199 | 3.276 | 8.2294 | 8.651 | 15.852 | 16.618 |

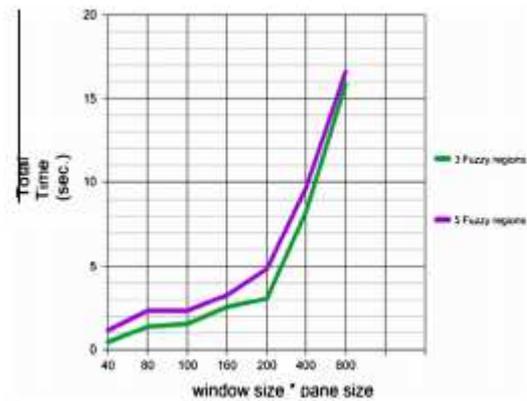
جدول 6

4-2 پیچیدگی حافظه

پیچیدگی حافظه با اندازه پنجره و ترتیب ایتام ها تر تراکنش جریان داده ها تغییر می کند. با فرض بدترین سناریو، تعداد گره ها در DFFP را می توان به صورت $O(|P| * |I|)$ در نظر گرفت.



شکل 6 الف: توزیع زمان اجرای تصادفات یا برخورد ها بر تغییرات مناطق فازی



شکل 6 ب: توزیع زمان اجرای خرده بر تغییرات مناطق فازی

5- نتایج آزمایشی

برای ارزیابی اعتبار روش پیشنهادی، آزمایشات با استفاده از مجموعه داده های واقعی انجام می شوند. دو مجموعه متفاوت از داده ها انتخاب می شوند مجموعه داده های تصادف و جزئی/ این را می توان از سایت fimi.cs.helsinki.fi/data دانلود کرده و در جدول 5 خلاصه سازی کرد.

5-1 پارامتر های مورد نیاز

دو پارامتر اصلی دارای اثر زیادی بر روی عملکرد نتایج آزمایشی، انتخاب اندازه پنجره و انتخاب تابع عضویت با مناطق مجموعه های فازی دارند.

5-1-1 اندازه پنجره و قاب

درخت DFFP به طور دینامیک خود را بعد از هر کشوی و باز شدن پنجره بازسازی می کند. زمان اجرا و پیچیدگی حافظه بر اساس پارامتر های پنجره تغییر می کند.

5-1-2 توابع عضویت و مجموعه های فازی

یک مرحله مهم در FFP_USTREAM، انتخاب توابع عضوی و تعداد مجموعه های فازی است. از این روی توابع عضویت مناسب باید بر اساس ماهیت مجموعه های داده و برنامه کاربر انتخاب شوند. در این مطالعه، جعبه ابزار منطق فازی در مطلب 7 استفاده می شود.

در شکل 4، تابع عضویت با سه منطقه فازی پایین، متوسط و بالا نشان داده شده است. تابع عضویت با 5 منطقه فازی (بسیار پایین، پایین، متوسط، بالا و بسیار بالا) در شکل 5 نشان داده شده است.

5-2 تحلیل نتایج

در بخش های زیر، نتایج ساختار درختی DFFP نشان داده شده است. تحلیل های آزمایشی گسترده نشان می دهند که درخت DFFP از حیث ذخیره حافظه زمانی کارآمد است که یافتن الگوهای دقیق فازی از جریان های داده های فراگیر پر سرعت بالا باشد

5-2-1 روند تغییرات زمان اجرا

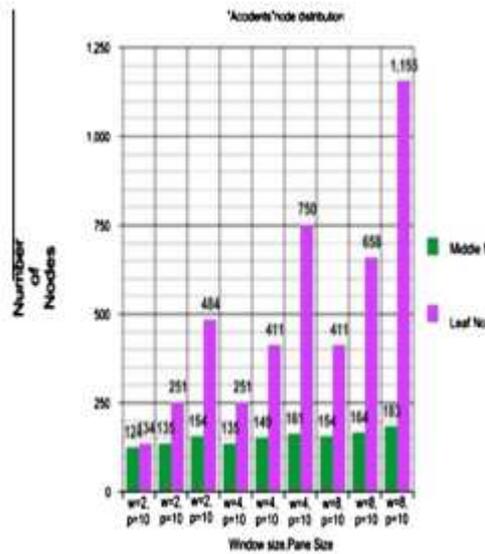
برای هر مجموعه داده ها، زمان اجرای کل درخت DFFP بر اساس تغییرات در مناطق مجموعه فازی ارزیابی شد. نتایج در جدول 6 و شکل 6 نشان داده شده است. در نمودار، محور های Y زمان کل را از جمله زمان ساخت درخت DFFP و زمان بازسازی درخت را نشان می دهند. محور ایکس تغییرات حاصل اندازه پنجره و اندازه قاب را نشان می دهد.

- قاب های بزرگ تر و اندازه پنجره های بزرگ تر نشان دهنده طولانی بودن زمان باز سازی کل درخت برای دو نوع مجموعه داده هستند
- تعداد مناطق مجموعه فازی بر زمان اجرای ساخت درخت اثر می گذارند. مدت زمان اجرای طولانی با مناطق با مجموعه های فازی لازم است.
- زمان اجرای طولانی تر برای مجموعه داده های متراکم در مقایسه با مجموعه داده های پراکنده نیاز است.

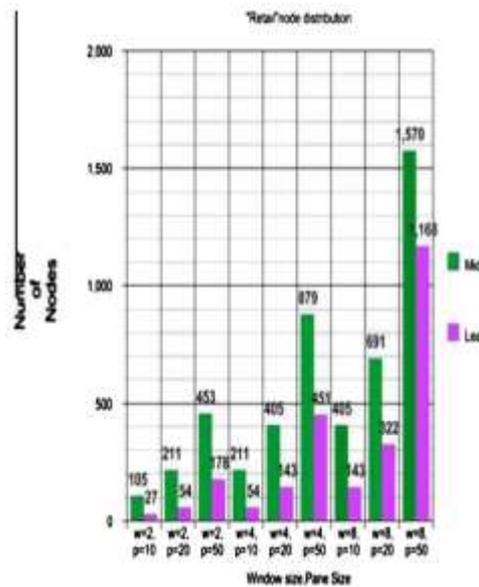
Table 7 Node distribution.

| Data sets | Window size | Pane size | | | | | | | | |
|-----------|-------------|-----------|------|-------|------|------|-------|------|------|-------|
| | | 10 | | | 20 | | | 50 | | |
| | | Mid. | Leaf | Total | Mid. | Leaf | Total | Mid. | Leaf | Total |
| Accident | 2 | 124 | 134 | 258 | 135 | 251 | 386 | 154 | 484 | 638 |
| | 4 | 135 | 251 | 386 | 149 | 411 | 560 | 164 | 750 | 914 |
| | 8 | 149 | 411 | 560 | 161 | 658 | 819 | 183 | 1155 | 1338 |
| Retail | 2 | 105 | 27 | 132 | 211 | 54 | 265 | 453 | 178 | 631 |
| | 4 | 211 | 54 | 265 | 405 | 143 | 548 | 879 | 451 | 1330 |
| | 8 | 405 | 143 | 548 | 691 | 322 | 1013 | 1570 | 1168 | 2738 |

جدول 7



شکل a7



شکل b7

5-2-2 کارایی حافظه

آزمایشاتی برای صحت سنجی و تعیین ملزومات حافظه برای ساخت درخت dffp بر روی مجموعه داده های مختلف با اندازه پنجره متغیر انجام شد. FFP_USTREAM موضوعات پنجره را به طور کامل برای کاهش اسکن دیتابیس به یک بار پوشش می دهد. از این روی، مقادیر آستانه بر حافظه مورد نیاز در این تحقیق اثری ندارد. تعداد گره ها در درخت DFFP با اندازه پنجره و تغییرات اندازه قاب در جدول 7 نشان داده شده است.

لازم به ذکر است که تعداد گره ها ثابت است و با تغییرات اندازه پنجره برای مجموعه داده های مختلف این میزان ثابت می کند.

| Data sets | Fuzzy frequent patterns | | | | |
|-------------|-------------------------|-------------|--------------|--------------|-------------|
| | 1-Itemsets | | 2-Itemsets | | |
| | Items | Count | First items | Second items | Count |
| Accidents | 1.Very High | 3 | 1.Very High | 2.Low | 3.8 |
| | 10.High | 4.6 | 7.Low | 5.High | 3.2 |
| | 24.Very High | 8 | 8.Very Low | 8.High | 3.2 |
| | 25.Very High | 9.8 | 9.Very High | 12.Medium | 9 |
| | 26.Very High | 9.8 | 13.Very High | 13.Medium | 9 |
| | 27.Very High | 11.2 | 15.Very High | 15.Medium | 9.4 |
| | 28.Very High | 2.2 | 17.Very High | 16.Medium | 9.4 |
| | 29.Very High | 2.6 | 43.Very High | 41.Medium | 7.4 |
| | Retail | 0.Very High | 1 | 1.Low | 0.Very High |
| 1.Very High | | 1 | 2.Low | 1.Very High | 0.6 |
| 10.Very Low | | 0.6 | 4.Low | 2.Very High | 0.6 |
| 2.Very High | | 4.8 | 5.High | 3.Very High | 0.6 |
| 22.Low | | 7.2 | 12.Medium | 4.Very High | 0.8 |
| 23.High | | 4 | 18.Very High | 5.Very High | 0.8 |
| 24.High | | 3.4 | 19.Very High | 39.Very High | 2.8 |
| | | | 20.Very High | 48.Very Low | 1.6 |
| | | | 21.Very High | 49.Very Low | 2 |

جدول 8

3-2-5 فهرست الگوهای تکراری فازی

بعد از ساخت درخت DFFP، الگوریتم قوانین رابطه فازی برای استخراج الگوهای فازی دقیق استفاده می شود. مجموعه داده های ذکر شده در جدول 5 و توابع عضویت با مجموعه 5 داده فازی (بسیار پایین، پایین، متوسط، بالا و بسیار بالا) در شکل 5 استفاده می شوند. جدول 8 نمونه ای از مجموعه ایتم های 1 و 2 را با مقیاس کل آن ها در همه تراکنش های محاسبه شده به صورت مقدار شمارش برای اندازه قاب=20 نشان می دهد. مجموعه های کامل برای مکان های بدون فضا نشان داده نشده است. پشتیبانی حداقل از پیش تعریف شده برای داده های تصادفی به صورت انتخاب شده=5٪ می باشد که برای داده های جزئی 2.5 درصد است. از این روی، قوانین ارتباط فازی را می توان در جدول 9 نشان داد.

توجه کنید که اعداد نوشته شده قبل از اصطلاح لینگویستیک نشان دهند شرایط آب و هوایی (آفتابی، ابری) برای مجموعه داده های تصادفی می باشند که ایتم های خریداری شده شیر، نان و تخم مرغ برای مجموعه داده های جزئی هستند.

| Data sets | Fuzzy association rules |
|-----------|---|
| Accidents | - The Very High value of item: 17 \Rightarrow Medium value of item: 16 |
| | - The Very High value of item: 9 \Rightarrow Medium value of item: 12 |
| | - The Very High value of item: 43 \Rightarrow Medium value of item: 41 |
| | - The Very Low value of item: 1 \Rightarrow Low value of item: 2 |
| Retail | - The Very High value of item: 19 \Rightarrow Very High value of item: 39 |
| | - The Very High value of item: 21 \Rightarrow Very Low value of item: 49 |
| | - The Very High value of item: 20 \Rightarrow Very Low value of item: 48 |
| | - The Low value of item: 1 \Rightarrow Very High value of item: 0 |
| | - The High value of item: 5 \Rightarrow Very High value of item: 3 |

جدول 9

6- نتیجه گیری و کار های آینده

بر اساس منابع بررسی شده، هیچ روشی در منابع وجود ندارد که نشان دهد قوانین کاوش ارتباط فازی از جریان های داده های فراگیر پرسرعت وجود دارند. در این مقاله، یک روش موثر برای کاربرد های فراگیر و خصوصیات مربوطه در بخش 2-3 نشان داده شده است. هم چنین، یک الگوی فازی دینامیک ترکیب یاز الگوی درختی فازی با مفهوم بازسازی درخت دینامیک است. روش پیشنهادی FFP_USTREAM در بسیاری از شرایط به مدیران کمک می کند تا تصمیمات مهم و انعطاف پذیر بگیرند. نظیر موارد زیر

- تعیین منابع مورد نیاز در برنامه ها و کاربرد های جزئی
- تعیین روش های درمان در پزشکی
- تعیین روش های احتیاطی در ایمنی جاده ها



این مقاله، از سری مقالات ترجمه شده رایگان سایت ترجمه فا میباشد که با فرمت PDF در اختیار شما عزیزان قرار گرفته است. در صورت تمایل میتوانید با کلیک بر روی دکمه های زیر از سایر مقالات نیز استفاده نمایید:

لیست مقالات ترجمه شده ✓

لیست مقالات ترجمه شده رایگان ✓

لیست جدیدترین مقالات انگلیسی ISI ✓

سایت ترجمه فا ؛ مرجع جدیدترین مقالات ترجمه شده از نشریات معتبر خارجی