اولین کنفرانس مهندسی فناوری اطلاعات مکانی
دانشگاه صنعتی خواجه نصیرالدین طوسی
دانشکده مهندسی نقشه‌برداری
۲۹ و ۳۰ دی ماه ۱۳۹۴

The 1st National Conference on
Geospatial Information Technology
K.N.Toosi University of Technology
Faculty of Geomatics Engineering
19 - 20 January 2016

استخراج و بهنگام‌رسانی خودکار اطلاعات توصیفی راه از داده‌های خط‌سیر با استفاده از ماشین بردار پشتیبان

رضا محمدی*[۱]، مهدی فرنقی[۲]

۱- . دانشجوی کارشناسی ارشد سیستم‌های اطلاعات مکانی دانشگاه خواجه‌نصیرالدین طوسی
۲. استادیار گروه سیستم‌های اطلاعات مکانی دانشگاه خواجه‌نصیرالدین طوسی

چکیده :

نقشه‌های رقومی بهنگام نقش بسزایی در ارائه خدمات باکیفیت در کاربردهایی نظیر ناوبری، توریسم، سرویس‌های نقشه تحت وب و سامانه‌های مکان‌مبنا ایفا می‌کنند. امروزه استفاده از داده‌های داوطلبانه، به‌منظور استخراج اطلاعات توصیفی راه‌ها موردتوجه قرارگرفته است. روش‌های ارائه‌شده به‌منظور استخراج نقشه رقومی راه‌ها با استفاده از داده‌های داوطلبانه، عموماً ایستا بوده و به تغییرات پیوسته شبکه راه‌ها که توسط داده‌های خط‌سیر کاربران گزارش می‌شوند توجه نمی‌کنند. ایستا بودن این روش‌های ارائه‌شده باعث می‌شود که بهنگام رسانی نقشه‌های راه داوطلبانه در بازه‌های زمانی قابل توجهی صورت گیرد. استفاده از عامل‌ها به دلیل قابلیت حس محیط و توانایی عکس‌العمل در برابر تغییرات محیط، تولید نقشه به‌صورت پویا را امکان‌پذیر می‌کند. این مقاله قصد دارد روشی پویا به‌منظور استخراج و بهنگام رسانی اطلاعات توصیفی راه توسط عامل‌های کاوشگر با استفاده از روش ماشین بردار پشتیبان ارائه دهد. به منظور گویاسازی راه-ها، عاملی توسعه داده شد که ابتدا با حرکت بر روی راه‌های دارای اطلاعات توصیفی و با در نظرگرفتن ویژگی‌های داده‌های خط‌سیر مربوط به این راه‌ها، آموزش دیده و سپس با حرکت بر روی راه‌های فاقد اطلاعات توصیفی و با در نظرگرفتن ویژگی‌های داده‌های خط‌سیر مربوط به این راه‌ها، به استخراج اطلاعات توصیفی نوع راه‌ها می‌پردازد. مقادیر دقت کلی و ضریب کاپا برای نتایج نهایی طبقه‌بندی روش پیشنهادی این مقاله به ترتیب برابر ۷۵/۵۵ و ۶۷ درصد حاصل شد.

واژه‌های کلیدی : داده‌های داوطلبانه، استخراج خودکار راه، عامل، عامل کاوشگر، گویاسازی

نویسنده مکاتبه کننده : رضا محمدی
آدرس پستی :
تلفن :
آدرس پست الکترونیک : mohammadireza@yahoo.com

۱- مقدمه

در بسیاری از کشورها، سازمان‌ها و ارگان‌های رسمی متولیان تولید اطلاعات مکانی می‌باشند. علی‌رغم هزینه و زمان قابل‌توجهی که همه‌ساله صرف تولید و بهنگام رسانی اطلاعات مکانی می‌شود، هنوز هم بازه زمانی بهنگام رسانی این داده‌ها طولانی و غیرقابل‌قبول به نظر می‌رسد. از طرفی امروزه تعداد استفاده‌کنندگان از شبکه‌های اجتماعی[1] رو به افزایش است. این شبکه‌ها بر پایه‌ی فناوری وب ۲.۰ و در تعامل پیوسته با کاربران ایجادشده‌اند. این تعامل منجر به ایجاد حجم عظیمی از داده‌ها در محیط وب شده است. در یک محیط داوطلبانه هر کاربر، به‌عنوان یک سنجنده[2] عمل کرده، داده‌های مکانی را جمع‌آوری و برای سرور مرکزی ارسال می‌کند [۱] و از این طریق منابع اطلاعات مکانی شبکه‌های اجتماعی به‌صورت پیوسته تغذیه و غنی‌سازی می‌شود. این داده‌های داوطلبانه می‌توانند به‌عنوان منبع اطلاعاتی مناسبی به‌منظور استخراج نقشه‌های راه مورداستفاده قرار گیرند. این داده‌ها همچنین به دلیل ماهیت پویایی آن‌ها، معمولاً بهنگام بوده و نسبت به داده‌های مکانی استاندارد بهتر می‌توانند تغییرات دنیای واقعی را نمایش دهند [۲].

در زمینه استخراج اطلاعات توصیفی راه‌ها با استفاده از داده‌های خط‌سیر،[۳] با فرض قابل‌قبول بودن تعداد و توزیع داده‌های خط‌سیر در هر راه، راه‌حلی تجربی به منظور تعیین عرض و تعداد لاین‌های راه ارائه دادند. [۴] برای تشخیص راه‌های مشابه در نقشه قدیمی و جدید و یا راه‌های اخیراً مسدود و یا ایجادشده از یک گراف وزن‌دار دوبخشی استفاده نمودند. [۵] با استفاده از اطلاعاتی نظیر سرعت، شتاب و جهت راه‌ها اطلاعات نوع راه‌ها را استخراج نمودند. [۶] با در نظر گرفتن ویژگی‌های سرعت، شتاب، جهت و امتداد راه، با استفاده از شبکه عصبی مصنوعی[3] روشی به منظور استخراج اطلاعات نوع راه‌ها ارائه نمودند.

روش‌های ارائه‌شده به‌منظور استخراج اطلاعات توصیفی راه‌ها با استفاده از داده‌های داوطلبانه، عموماً ایستا بوده و به تغییرات پیوسته شبکه راه‌ها که توسط داده‌های خط‌سیر کاربران گزارش می‌شوند توجه نمی‌کنند. ایستا بودن روش‌های ارائه‌شده به‌منظور استخراج اطلاعات توصیفی راه‌ها باعث می‌شود که بهنگام رسانی این اطلاعات به‌طور بهینه صورت نگیرد. استفاده از عامل‌ها به دلیل قابلیت حس محیط و توانایی عکس‌العمل در برابر تغییرات محیط، استخراج اطلاعات توصیفی راه‌ها به‌صورت پویا را امکان‌پذیر می‌کند. این مقاله قصد دارد روشی پویا به‌منظور استخراج و بهنگام رسانی اطلاعات توصیفی راه توسط عامل‌های کاوشگر با استفاده از روش ماشین بردار پشتیبان ارائه دهد. رویکرد این مقاله بدین گونه است که با استفاده از اطلاعات مشارکتی جمع‌آوری‌شده توسط کاربران، راه‌های فاقد اطلاعات توصیفی به‌طور خودکار و پویا شناسایی سپس طبقه‌بندی و درنهایت بهنگام خواهند شد.

ساختار مقاله در ادامه به این شرح است: در بخش دوم راه‌کار ارائه‌شده در این مقاله به‌منظور استخراج اطلاعات توصیفی راه توسط عامل گویاساز به‌صورت کامل تشریح شده است. بخش سوم به پیاده‌سازی و تحلیل و ارزیابی نتایج طبقه‌بندی می‌پردازد. نتایج و پیشنهادات نیز بخش پایانی این مقاله را شکل داده است.

۲-مواد و روش‌ها

همان‌گونه که گفته شد، روش‌های ارائه‌شده درزمینه استخراج راه از داده‌های خط‌سیر که در این مطالعه مورد بررسی قرار گرفت عموماً ایستا بوده و به تغییرات پیوسته شبکه راه‌ها که توسط داده‌های خط‌سیر کاربران گزارش می‌شوند توجه نمی‌کنند. این روش‌ها معمولاً یک‌بار روی‌داده‌های خط اجراشده و شبکه‌ی راه را استخراج می‌نمایند. ایستا بودن روش‌های استخراج راه از داده‌های خط‌سیر منجر به عدم بهنگام رسانی نقشه‌های شبکه راه متناسب با تغییرات دنیای

[1] Social networks
[2] Sensor
[3] Artificial Neutral Network

واقعی می‌گردد. استفاده از عامل‌ها به‌عنوان مولفه‌های نرم‌افزاری باقابلیت عکس‌العمل نسبت به تغییرات محیط، به‌عنوان ابزار بهنگام رسانی نقشه‌های راه‌ها، می‌تواند سازوکارهای ایستایی ارائه‌شده را به سازوکارهای پویا تبدیل نماید.

راسل [۷] عامل‌ها را کدهای کامپیوتری معرفی می‌کند که از قابلیت درک محیط و عکس‌العمل در برابر تغییرات برخوردار می‌باشند. نوع ویژه‌ای از عامل‌های هوشمند که به‌منظور شناسایی و نمایه‌گذاری[۴] صفحات مفید وب توسط موتورهای جست‌وجو[۵] مورد استفاده قرار می‌گیرند، عامل‌های کاوشگرهای وب[۶] نامیده می‌شوند [۸]. عامل‌های کاوشگر وب به‌طور پیوسته و خودکار بر روی صفحات وب (که با لینک به یکدیگر متصل شده‌اند) حرکت نموده و به کاوش محتوی صفحات وب می‌پردازند.

این مقاله بخشی از یک پژوهش به‌منظور توسعه یک سیستم چند عامله[۷] است. یکی از عامل‌های توسعه داده‌شده در این تحقیق، عامل گویاساز است. در این بخش از پژوهش عاملی به‌منظور استخراج ویژگی توصیفی نوع راه برای راه‌های فاقد این اطلاعات توسعه داده خواهد شد. شکل ۱ محیط عمل عامل گویاساز شامل نقشه مبنا[۸] بهمراه فضای داده‌های داوطلبانه خط‌سیر را نمایش می‌دهد. این عامل ابتدا آموزش دیده و سپس به استخراج اطلاعات توصیفی راه‌های فاقد این اطلاعات می‌پردازد. به‌منظور آموزش عامل گویاساز، این عامل بر روی داده‌های خط‌سیر متعلق به راه‌های دارای اطلاعات توصیفی موجود در نقشه مبنا حرکت نموده و با در نظر گرفتن برچسب[۹] نوع راه که از نقشه مبنا موجود است و همچنین ویژگی‌های[۱۰] (نظیر تعداد، تراکم، سرعت و شتاب) داده‌های داوطلبانه جمع‌آوری‌شده، با استفاده از روش ماشین بردار پشتیبان آموزش داده می‌شود. پس از آموزش عامل گویاساز این عامل در فضای نقشه مبنا حرکت نموده و به کشف راه‌های فاقد اطلاعات توصیفی می‌پردازند. سپس با حرکت بر روی داده‌های خط‌سیر متعلق به راه‌های فاقد اطلاعات توصیفی به استخراج این اطلاعات می پردازد.

شکل ۱- نقشه مبنا به‌همراه فضای داده‌های خط‌سیر

[4] Indexing
[5] Search engine
[6] Web Crawler-Agents
[7] Multi-Agent System
[8] Ground truth map
[9] Lable
[10] Features

۲-۱-گویاسازی

۲-۱-۱-آموزش عامل گویاساز

شبه کد ۱ نحوه عمل عامل گویاساز را تشریح می‌کند. در ابتدا به‌منظور آموزش عامل گویاساز، این عامل بر روی راه‌های دارای اطلاعات توصیفی از نقشه مبنا حرکت می‌کند (خط ۱ شبه کد). سپس به تعیین ویژگی‌های داده‌های خط‌سیر مربوط به این راه‌ها می پردازد.

```
1: The Information Extractor Agent goes on one road from the base map:
2: The road divide into segments
3: For each segments do
4:        Make a query to determine data trajectories that are belonging to the segment
5:        Features (i) ◄——— Calculate features for the data trajectories
6:        If (The road has descriptive information) do
7:                Training the Information Extractor Agent:
8:                Label (i) ◄——— Get the label of the road from the base map
9:                Train SVM: svm (i) (Features (i), Label (i))
10:       Else if (The road has not descriptive information) do
11:               Classification by the Information Extractor Agent:
12:               Label(i) ◄——— Classification by SVM (Features)
13:       End if
14: End for
15: Road's Label = Mode (Label(i))
```

شبه کد ۱: نحوه عمل عامل گویاساز

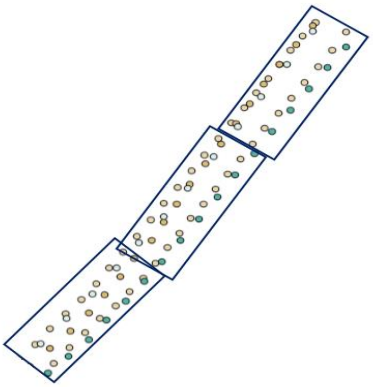
- **تعیین ویژگی‌های داده‌های خط‌سیر**

از آنجایکه توزیع داده‌های خط‌سیر در سرتاسر راه همگون نمی‌باشد [۹]، لذا در ابتدا به‌منظور آموزش بهتر عامل گویاساز، راه به بخش[11]هایی تقسیم خواهد شد (خط ۲ شبه کد). سپس به‌منظور تعیین نقاط خط‌سیر متعلق به هر بخش راه از یک پرسش و پاسخ مکانی استفاده شد (خط ۴ شبه کد). بدین منظور، بجای استفاده از یک پرسش و پاسخ به شکل دایره، از پرسش و پاسخ مستطیلی شکل استفاده شد(شکل ۳). این‌کار تا حدودی از وارد شدن داده‌های خط‌سیر غیر مرتبط (داده‌های خط‌سیر متعلق به راه‌های دیگر) جلوگیری می‌کند. بااین‌حال، در مکان‌هاییکه که دو راه فاصله کمی تا یکدیگر دارند (مانند بزرگراه‌ها، چهارراه‌ها و تقاطع‌ها) این پرسش و پاسخ ممکن است شامل داده‌های خط‌سیر متعلق به راه‌های دیگر نیز باشند. به منظور جلوگیری از این مشکل داده‌های خط‌سیر متعلق به کاربران مختلف بر اساس شناسه[12] کاربران از یکدیگر جدا خواهند شد.

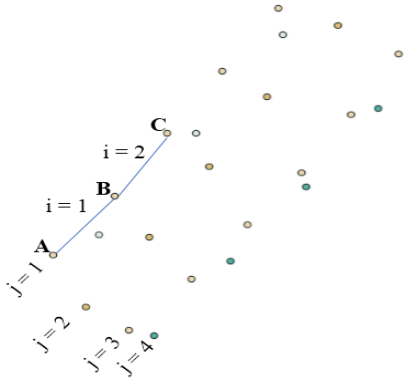
پس از تعیین داده‌های خط‌سیر مربوط به هر بخش، این داده‌ها به‌منظور محاسبه ویژگی‌های سرعت، شتاب، تعداد و تراکم مورداستفاده قرار خواهند گرفت (خط ۵ شبه کد). به منظور تعیین ویژگی‌های سرعت و شتاب داده‌های خط‌سیر،

[11] Segment
[12] Id

با توجه به اینکه محاسبه شتاب برای حداقل سه نقطه متوالی امکان‌پذیر است، لـذا ابعـاد هـر یـک از بخـش‌هـا طـوری انتخاب خواهد شد که در هر بخش از هر خط‌سیر (که بر اساس شناسه کاربران از یکدیگر جدا شده‌اند) حداقل ۳ نقطـه متوالی موجود باشد. از طرفی طول این بخش نباید بیشتر از یک‌چهارم طول مسیر باشند. بیشترین ابعـاد بخـش‌هـا بـر اساس سعی و خطا تعیین شد. شکل ۳ بخش‌های ایجادشده برای یک‌راه را نمایش می‌دهد. که در آن داده‌های خط‌سیر متعلق به هر کاربر با رنگ‌های مختلفی نمایش داده‌شده‌اند.



شکل ۳- پرسش و پاسخ مکانی استفاده شده در هر بخش به‌منظور تعیین خط‌سیرهای مرتبط به راه



شکل ۴- داده‌های خط‌سیر متعلق به یک بخش

شکل ۴ داده‌های خط‌سیر متعلق به یک بخش را نمایش می‌دهد. با در نظر گرفتن شکل ۴، سـرعت و شـتاب داده‌هـای خط‌سیر از روابط (۲-۸) و (۲-۹) قابل محاسبه است.

$$v_i = \Delta x_{(z,z+1)} \Big/ \Delta t_{(z,z+1)} \qquad (۲-۸)$$

$$a_i = \Delta v_{(i,i+1)} \Big/ \Delta t_{(i,i+1)} \qquad (۲-۹)$$

که در آن v_i سرعت محاسبه شده برای دو نقطه متوالی $(z, z+1)$ و a_i شتاب محاسبه شده برای دو اتصـال متـوالی $(i, i+1)$ از یک خط‌سیر است. پس از محاسبه سرعت و شتاب برای اتصالات متوالی، سرعت و شتاب متوسط برای هر خط-سیر از روابط (۲-۱۰) و (۲-۱۱) و در نهایـت سـرعت و شـتاب متوسـط بـرای هـر بخـش از روابـط (۲-۱۲) و (۲-۱۳) محاسبه خواهند شد.

$$v_j = \sum_{i=1}^{n} v_i \Big/ n \qquad (۲-۱۰)$$

$$a_j = \sum_{i=1}^{n} a_i \Big/ n \qquad (۲-۱۱)$$

$$\overline{v_k} = \sum_{j=1}^{m} v_j \Big/ m \qquad (۲-۱۲)$$

$$\overline{a_k} = \sum_{j=1}^{m} a_j \Big/ m \qquad (۲-۱۳)$$

که در آن $\overline{v_k}$ و $\overline{a_k}$ به ترتیب سرعت و شتاب متوسط محاسبه شده و n تعداد اتصـالات موجـود از یـک خـط‌سـیر و m تعداد خط‌سیرهای واقع شده در بخش kام است.

تعداد و توزیع داده‌های خط‌سیر موجود در هر بخش به‌عنوان دیگر ویژگی تاثیرگذار در تعیین استخراج اطلاعات توصیفی راه‌ها در نظر گرفته شد. با در نظر گرفتن ویژگی‌های محاسبه شده برای هر بخش و از طرفی معلوم بودن برچسب نوع راه از نقشه مبنا (خط ۸ شبه کد)، جدول توصیفات برای هر راه تشکیل خواهد شد. جدول ۲-۱، ویژگی-های محاسبه شده برای هر بخش و همچنین برچسب راه برای دو راه مختلف را (به‌عنوان نمونه) نمایش می‌دهد.

جدول ۲-۱: نمونه ویژگی‌های تعیین‌شده داده‌های خط‌سیر متعلق به راه‌ها به‌منظور آموزش و تست روش SVM

شناسه راه	شناسه بخش	تعداد داده‌های خط سیر	تراکم	سرعت میانگین (متر/ثانیه)	شتاب میانگین (متر/مجذور ثانیه)	برچسب نوع راه
۱	۱	۱۲۲	۳۴۸۰.۵۰۲	۱۵.۰۵۷	۰.۰۵۹۱	بزرگراه
	۲	۱۶۸	۵۲۷۷.۶۶۱	۱۰.۸۰۳	۰.۰۵۵	بزرگراه
	۳	۷۳	۳۰۱۶.۱۶۹	۲۰.۳۳۲	۰.۰۷۲	بزرگراه

- **آموزش عامل گویاساز**

به‌منظور آموزش عامل گویاساز از یک روش طبقه‌بندی در فضای داده‌های خط‌سیر استفاده شد. روش‌های طبقه‌بندی مختلفی وجود دارد که از این میان روش svm مورد استفاده قرار گرفت. روش svm مدل‌های یادگیری غیر پارامتریک نظارت‌شده‌ای هستند که با الگوریتم‌های مرتبط و آنالیز داده، الگوهای داده‌ها را تشخیص می‌دهند. این روش در مقایسه با سایر روش‌های طبقه‌بندی نیاز به داده‌های آموزشی کمتری دارد. همچنین دقت این روش در مقایسه با دیگر روش‌های طبقه‌بندی بیشتر است [۱۰]. هدف svm یافتن ابرصفحه[۱۳] بهینه‌ای است که دو کلاس را به‌گونه‌ای از هم جدا کند که فاصله نزدیک‌ترین نمونه‌های آموزشی دو کلاس به این ابرصفحه بیشینه گردد. ۳درشرایطی که از این روش به منظور طبقه بندی کلاس هایی با تعداد بیشتر مد نظر باشد، استراتژی‌های مختلفی قابل اتخاذ است. از آن جایی‌که در این مقاله هدف طبقه‌بندی راه‌ها به چهار کلاس است، لذا استراتژی "یک در برابر یک[۱۴]" مورداستفاده قرار گرفت [۱۱]. در استراتژی مذکور برای هر جفت از کلاس‌های ممکن طبقه‌بندی جداگانه انجام می‌گیرد. پس از تشکیل جدول توصیفات برای هر بخش، ویژگی‌ها و برچسب‌های تعیین‌شده به‌منظور آموزش عامل گویاساز با استفاده از روش svm مورداستفاده قرار خواهند گرفت (خط ۹ شبه کد).

۲-۱-۲-استخراج اطلاعات توصیفی

پس از آموزش عامل گویاساز، به‌منظور استخراج اطلاعات توصیفی راه، این عامل بر روی راه‌های فاقد اطلاعات توصیفی حرکت می‌کند. مشابه استراتژی که در آموزش عامل استخراج‌کننده پی گرفته شد، این عامل هر راه را به بخش‌هایی تقسیم نموده (خط ۲ شبه کد) سپس برای هر یک از این بخش‌ها با استفاده از یک پرسش و پاسخ مکانی، داده‌های خط‌سیر متعلق به این بخش تعیین خواهند شد (خط ۴ شبه کد). درنهایت نقاط خط‌سیر به‌دست‌آمده، به‌منظور محاسبه ویژگی‌های سرعت متوسط، شتاب متوسط، تعداد و تراکم، مورداستفاده قرار خواهند گرفت (خط ۵ شبه کد).

پس از تعیین جدول توصیفات برای بخش‌های مختلف یک‌راه، بر اساس ویژگی‌های محاسبه شده، کلاس نوع راه توسط عامل گویاساز، تعیین خواهد شد (خط ۱۲ شبه کد). مراحل تعیین برچسب نوع راه، برای هر بخش از یک‌راه تکرار خواهد شد. درنهایت از بین برچسب‌های استخراج‌شده، برچسبی با فراوانی بیشتر به‌عنوان برچسب نوع راه در نظر گرفته خواهد شد (خط ۱۵ شبه کد).

[13] Hyperplane
[14] one-against-one

۳-پیاده‌سازی و تحلیل نتایج

در این مقاله عامل گویاساز با استفاده از زبان برنامه‌نویسی جاوا توسعه داده شد. به دلیل یکپارچگی بالای جاوا با پایگاه داده متن‌باز PostgreeSQL از این نرم‌افزار و از ماژول PostGIS به‌منظور ذخیره‌سازی داده‌های مکانی و اجرای پرسش و پاسخ‌های مکانی و توصیفی موردنیاز، استفاده شد. همچنین به‌منظورآموزش عامل گویاساز از کتابخانه LibSVM به‌همراه کتابخانه weka استفاده شد. کتابخانه‌های متن‌باز Weka[۱۵] و LibSVM[۱۶] تحت مجوز GPL [۱۷] منتشرشده‌اند که دارای مجموعه‌ای از الگوریتم‌هایی مختلف طبقه‌بندی در زمینه یادگیری ماشین[۱۸] می‌باشند. علاوه بر این امکان استفاده از این کتابخانه‌ها به‌منظور طبقه‌بندی چند کلاسه نیز فراهم شده است [۱۲].

۳-۱-داده‌های مورداستفاده و پیش‌پردازش داده‌ها

داده‌های مورداستفاده در این مقاله شامل مجموعه داده پروژه GeoLife و نیز نقشه مبنای شهر پکن است. این مجموعه داده توسط ۱۸۲ کاربر در مدت سه سال (آوریل 2007 تا اوت2012) جمع‌آوری‌شده است. هر خط‌سیر در این مجموعه داده با استفاده از مجموعه‌ای از نقاط بافاصله زمانی مشخص، نمایش داده می‌شود که شامل طول و عرض جغرافیایی، ارتفاع و زمان نمونه‌برداری (x,y,z,timestamp) است. [۱۳]. همچنین از نقشه ارائه‌شده توسط OSM[۱۹] به‌عنوان نقشه مبنا استفاده شد. با توجه به نقشه مبنای موجود راه‌های شهری در چین به چهار کلاس بزرگراه، راه اصلی، راه فرعی و راه محلی[۲۰] تقسیم می‌شوند [۱۴].

نویز ذاتی سیستم تعیین موقعیت جهانی[۲۱] در دقت و صحت داده‌های جمع‌آوری‌شده تأثیرگذار است. در شرایط خاصی این نویز می‌تواند منجر به ایجاد راه‌های جعلی گردد. علاوه بر این، داده‌های خط‌سیر زمانی می‌توانند بیانگر راه باشند که توسط خودروهای در حال حرکت جمع‌آوری‌شده باشند [۱۵]. همچنین ازآنجایی‌که داده‌های جمع‌آوری‌شده تشکیل‌دهنده یک خط‌سیر، بر اساس توالی زمانی به یکدیگر متصل می‌شوند، لذا در برخی از موارد یک اتصال غیرمنطقی بین سفرهای[۲۲] مختلف وجود دارد. بنابراین به‌منظور استفاده از این داده‌ها باید این سفرها از یکدیگر جدا شوند [۱۶]. ازآنجایی‌که هدف ارائه راهکاری باقابلیت کار با داده‌های با حجم بالا است، در ابتدا پالونه‌ای، با در نظر گرفتن قوانین زیر به‌منظور حذف داده‌های اضافی و عملکرد بهتر سیستم، اعمال می‌گردد.

- قانون ۱: داده‌هایی با سرعت کمتر از ۱۰ کیلومتر بر ساعت حذف خواهند شد [۱۵].
- قانون ۲: اگر فاصله بین دونقطه بیشتر از ۳۰۰ متر و یا اینکه تغییر جهت[۲۳] آن‌ها بیشتر از ۴۵ درجه باشد، داده‌های خط‌سیر دو سفر مجزا در نظر گرفته می‌شوند [۱۶]. سپس سفرهایی با تعداد نقاط کمتر از ۲۰ داده در یک خط‌سیر از مجموعه داده حذف خواهند شد.

روش‌های زیادی به‌منظور کاهش نویز داده‌های خط‌سیر وجود دارند که می‌توان روش‌های: میانگین، میانه و پالونه کالمن[۲۴] و پالونه Particle را نام برد. به‌منظور استخراج ویژگی‌های سرعت و شتاب از داده‌های خط‌سیر کاهش نویز

---

[15] https://weka.wikispaces.com/
[16] https://www.csie.ntu.edu.tw/~cjlin/libsvm/
[17] General Public Licence
[18] Machine Learning
[19] Open Street Map
[20] Local road
[21] Global Positioning System (GPS)
[22] Trip: هر یک از فعالیت های داوطلب (کاری، ورزشی، تفریحی و ...) یک سفر نامیده می شود.
[23] Direction
[24] Kalman

ذاتی داده‌ها ضروری است. از بین روش‌های بیان‌شده به‌منظور کاهش نویز داده‌های خط‌سیر، روش‌های پالونه کالمن و Particle با در نظر گرفتن ویژگی پویایی داده‌های خط‌سیر در مدل‌سازی، با کاهش نویز ذاتی داده‌ها امکان دستیابی ویژگی‌های سرعت و شتاب از داده‌های داوطلبانه به طور صحیح و قابل اعتماد را فراهم می‌کنند [۱۷]. در این مقاله از پالونه کالمن به‌منظور کاهش نویز داده‌های خط‌سیر استفاده شد. سپس داده‌های باقیمانده به‌منظور استخراج راه توسط عامل‌ها مورداستفاده قرار گرفت. در ادامه نتایج استخراج راه بر اساس راه‌کار ارائه‌شده در این مقاله ارائه‌شده است.

جدول ۳ فراوانی داده‌های آموزشی و تست به تفکیک هر کلاس در نقشه مبنای موجود را نشان می‌دهد. که در آن راه‌هایی از نقشه مبنا که فاقد داده‌های داوطلبانه در دیتاست موجود هستند، حذف شده‌اند. از بین راه‌های موجود در نقشه مبنا، ۶۰ درصد راه‌ها به‌عنوان داده‌های آموزشی و ۴۰ درصد به‌عنوان داده‌های تست در نظر گرفته شد.

جدول ۳: فراوانی راه‌ها در نقشه مبنا

نوع کلاس / نوع داده	بزرگراه	راه اصلی	راه فرعی	راه محلی
آموزشی	۳۲	۵۳	۶۶	۳۴
تست	۲۰	۳۵	۴۴	۲۲

۳-۲-ارزیابی و تحلیل نتایج

در این مقاله به‌منظور ارزیابی و محاسبه دقت روش ارائه‌شده به‌منظور استخراج اطلاعات توصیفی راه‌ها از داده‌های خط‌سیر، پس از تشکیل ماتریس ابهام، از معیارهای دقت کل[۲۵] و ضریب کاپا[۲۶] استفاده شد. بدین منظور با مقایسه نتایج طبقه‌بندی عامل گویاساز با داده‌های تست، دقت نتایج طبقه‌بندی با استفاده از الگوریتم ارائه‌شده مورد ارزیابی قرار گرفت.

جدول ۴ نتایج ماتریس ابهام طبقه‌بندی برای چهار کلاس آورده شده است. عناصر روی قطر اصلی کلاس‌هایی که درست طبقه‌بندی شده‌اند و بقیه عناصر کلاس‌هایی که به اشتباه طبقه‌بندی شده‌اند را نشان می‌دهد.

جدول ۴: ماتریس ابهام طبقه‌بندی

پیش‌بینی / واقعی	بزرگراه	راه اصلی	راه فرعی	راه محلی
بزرگراه	۷۱/۴۲	۱۴/۲۸	۹/۵۲	۴/۷۸
راه اصلی	۲/۸۰	۸۳/۳۵	۸/۳۱	۵/۵۴
راه فرعی	۶/۸۵	۱۳/۳۶	۷۳/۳۳	۶/۴۶
راه محلی	۳/۴۷	۱۰	۱۶/۵۳	۷۰

جدول ۵: دقت کل هر کلاس برای خروجی طبقه‌بندی

نوع کلاس	دقت کلاس (درصد)
بزرگراه	۷۱/۴۲
راه اصلی	۸۳/۳۵
راه فرعی	۷۳/۳۳
راه محلی	۷۰
دقت کل	۷۵/۵۵
ضریب کاپا	۶۷

جدول ۵ نتیجه دقت طبقه‌بندی هر کلاس را نمایش می‌دهد. در شرایطی که تعداد داده‌های خط‌سیر گذرنده از یک راه به اندازه کافی نباشد، این داده‌ها از توزیع گوسی پیروی نمی‌کنند و به درستی مدلسازی نخواهند شد [۹]. لذا همانطور که انتظار می‌رود، از آنجاییکه تعداد داده‌های آموزشی و همچنین تعداد داده‌های خط‌سیر برای کلاس راه‌های اصلی نسبت به سایر کلاس‌ها بیشتر است، لذا دقت طبقه‌بندی این کلاس راه از سایر کلاس‌ها بیشتر است. همانطور که انتظار می‌رود، به دلیل کم بودن تعداد داده‌های آموزشی و همچنین داده‌های خط‌سیر جمع‌آوری شده برای کلاس راه-

[25] Overall Accuracy
[26] Kappa

های محلی، پایین‌ترین دقت طبقه‌بندی به این کلاس اختصاص یافته است. همچنین دقت کل و ضریب کاپا این طبقه‌بندی به ترتیب برابر با ۷۵/۵۵ و ۶۶ (درصد) شد.

۴-نتیجه‌گیری و پیشنهادات

ایستا بودن روش‌های ارائه‌شده به‌منظور استخراج راه از داده‌های داوطلبانه باعث می‌شود که علیرغم ماهیت پویایی داده‌های داوطلبانه که بیانگر تغییرات محیط هستند، عمل بهنگام رسانی به‌صورت گسسته و در بازه‌های زمانی قابل‌توجه صورت گیرد. در این مقاله با الهام از نحوه عملکرد عامل‌های کاوشگر وب در محیط وب، توسعه عاملی به‌منظور استخراج و بهنگام‌رسانی اطلاعات توصیفی نقشه‌های راه داوطلبانه با استفاده از روش طبقه‌بندی ماشین بردار پشتیبان در دستور کار قرار گرفت. استفاده از عامل‌ها به‌عنوان مولفه‌های نرم‌افزاری باقابلیت حس محیط و عکس‌العمل در برابر تغییرات لحظه‌ای محیط، می‌تواند سازوکار ایستایی ارائه‌شده را به سازوکارهای پویا تبدیل نماید. از آنجاییکه دقت طبقه‌بندی بطور مستقیم به تعداد داده‌های خط‌سیر متعلق به راه و نیز فراوانی کلاس راه در نقشه مبنا وابسته است، لذا کلاس راه‌اصلی با دقت ۸۳/۳۵ بهترین طبقه‌بندی را بخود اختصاص داده است. همچنین مقادیر دقت کلی و ضریب کاپا برای نتایج نهایی طبقه‌بندی روش پیشنهادی این مقاله به ترتیب برابر ۷۵/۵۵ و ۶۷ درصد حاصل شد.

مراجع

1- Goodchild, M.F., Citizens as sensors: the world of volunteered geography. GeoJournal, 2007. 69(4): p. 211-221.
2- N. Mohammadi, “Improving Logical Consistency in Volunteered Geospatial Information Environments”, Ph.D Thesis, K.N.Toosi University of Technology Faculty of Geodesy and Geomatics, 2014.
3- Jang, S ,.T. Kim, and S. Lee. Map generation system with lightweight GPS trace data. in Advanced Communication Technology (ICACT), 2010 The 12th International Conference on. 2010. IEEE.
4- Davics, J., A.R. Beresford, and A. Hopper, Scalable, distributed, real-time map generation. Pervasive Computing, IEEE, 2006. 5(4): p. 47-54.
5- Niehoefer, B., et al. GPS community map generation for enhanced routing methods based on trace-collection by mobile phones. in Advances in Satellite and Space Communications, 2009. SPACOMM 2009. First International Conference on. 2009. IEEE.
6- Ekpenyong, F., D. Palmer-Brown, and A. Brimicombe, Extracting road information from recorded GPS data using snap-drift neural network. Neurocomputing, 2009. 73(1): p. 24-36.
7- Russell, S. and P. Norvig, Artificial intelligence: a modern approach. 1995.
8- Dhenakaran, S. and K.T. Sambanthan, Web crawler-an overview. International Journal of Computer Science and Communication, 2011. 2(1): p. 265-267.
9- Zhang, L., F. Thiemann, and M. Sester. Integration of GPS traces with road map. in Proceedings of the second international workshop on computational transportation science. 2010. ACM.
10- Mountrakis, G., J. Im, and C. Ogole, Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 2011. 66(3): p. 247-259.
11- Yu, H. and S. Kim, SVM Tutorial—Classification, Regression and Ranking, in Handbook of Natural Computing. 2012, Springer. p. 479-506.
12- LibSVM. accessed 10 may 2015; Available from: https://www.csie.ntu.edu.tw/~cjlin/libsvm./
13- Trajectories, G.G., accessed 5 Jan 2015: p. Available from: http://research.microsoft.com/
14- Li, J., et al., Mining trajectory data and geotagged data in social media for road map inference. Transactions in GIS, 2015. 19(1): p. 1-18.

15- Li, J., et al., Integrated use of spatial and semantic relationships for extracting road networks from floating car data. International Journal of Applied Earth Observation and Geoinformation, 2012. 19: p. 238-247.

16- Zhang, D., et al. iBAT: detecting anomalous taxi trajectories from GPS traces. in Proceedings of the 13th international conference on Ubiquitous computing. 2011. ACM.

17- Zheng, Y. and X. Zhou, Computing with spatial trajectories. 2011: Springer Science & Business Media

Automatic road information extraction from trajectory data: An agent based approach

KEY WORDS: Agent, road Information extraction, Trajectory data, VGI, SVM.

ABSTRACT:

Up-to-date digital maps play a significant role as required data in various areas such as navigation, tourism, web-based map services and location-based services. Today, using volunteered trajectory data to extract road descriptive information is seriously considered by researcher in spatial data community. Previous studies about automatic extraction of road information from volunteered trajectory data are generally static. They have not considered the continuous changes of the road network which are reported dynamically by the volunteered trajectory data of users. Considering the fact that agents are capable of sensing and reacting to changes in the environment, it is possible to dynamically update road descriptive information based on the latest changes presented by the most recent trajectory data. The key idea of this study is to design and develop computer agents that dynamically search and update road descriptive information from volunteered trajectory data. To do so, armed with SVM, the information extractor agent moves the roads that have precise descriptive information (this type of roads are considered as a label) and determine parameters of relevant trajectory data. These descriptive information and label are used to train the SVM algorithm. It moves on the base map space and tries finding the roads without the descriptive information, and then finds the trajectory data belonging to the roads lacking the descriptive information, and calculates its features as inputs of the SVM classifier. Having these features, SVM classifier determines the type of the roads. These computer agents utilize SVM[27] to extract road descriptive information from trajectory data. The overall accuracy and kappa coefficient for the experimental results classification are respectively 75.55 and 67 percentage.

Keywords: Agent, road Information extraction, Trajectory data, VGI, SVM

[27] Support Vector Machine