

# TRAJECTORY OUTLIER DETECTION IN MOVING OBJECTS DATA

**Martin Pešek**

Master Degree Programme (3), FIT BUT

E-mail: xpesek07@stud.fit.vutbr.cz

Supervised by: Jaroslav Zendulka

E-mail: zendulka@fit.vutbr.cz

**Abstract:** Mining in trajectories of moving objects is currently a rapidly evolving area of research. This paper deals with the possibilities of the trajectory outlier detection with emphasis on the method, named TOP-EYE, which is designed to effectively identify evolving trajectory outliers at early stage with low false alarm rate.

**Keywords:** data mining, spatio-temporal data, moving objects, trajectory, outlier detection

## 1 ÚVOD

Získávání znalostí z databází je v posledních letech velmi rychle se rozvíjející oblast, která obzvláště vzhledem k obrovskému objemu různých dat nabývá stále více na užitečnosti a důležitosti. Jádrem procesu získávání znalostí je fáze dolování z dat, při které je aplikován konkrétní algoritmus za účelem extrakce požadovaných vzorů. Jednou z častých aplikací dolování z dat je analýza odlehlých hodnot. Jako odlehlou hodnotu lze chápat objekt, který se neshoduje s obecným chováním nebo modelem dat.

Zejména s rozvojem lokalizačních či dohledových systémů a senzorových sítí v nedávné době začal prudce narůstat objem ukládaných časoprostorových dat, čímž výrazně stoupá jejich potenciál. Nejčastěji se jedná o data pohybujících se objektů v podobě jejich trajektorií. Kromě rozvoje technologií pro modelování, dotazování a indexování přináší získávání znalostí z dat pohybujících se objektů, hlavně vzhledem k jejich složitosti, mnohé výzvy pro současný výzkum.

Mezi významné aplikace dolování v trajektoriích pohybujících se objektů patří dolování frekvencovaných pohybů, dolování periodických vzorů, klasifikace, predikce, shluková analýza a detekce anomálií či odlehlých trajektorií, tedy takových, které reprezentují neobvyklé chování.

Obsahem tohoto příspěvku je popis některých existujících přístupů k dolování odlehlých trajektorií se zaměřením na rozbor a implementaci metody nazvané TOP-EYE, která umožňuje efektivně identifikovat odlehlé trajektorie již v rané fázi jejich vývoje s nízkou mírou planých poplachů [1].

## 2 DETEKCE ODLEHLÝCH TRAJEKTORIÍ

Automatická detekce neobvyklého nebo podezřelého chování pohybujících se objektů je významnou úlohou v oblasti analýzy časoprostorových dat. Nejčastěji se jedná o hledání *odlehlých trajektorií*, tedy takových pohybů, které se výrazně odlišují od většiny ostatních. Každá trajektorie je obvykle reprezentována jako sekvence prostorových pozic pohybujícího se objektu v čase.

Pro dolování odlehlých trajektorií je možné zvolit různé přístupy. Jedním z nich je algoritmus nazvaný **TRAOD** [2]. Tento algoritmus sestává ze dvou fází. V první fázi je každá trajektorie rozdělena na množinu sub-trajektorií, přičemž za účelem zajištění kvality a efektivity je dělení prováděno nejprve hrubým způsobem a následně pak patřičně zjemněno. V druhé fázi algoritmu probíhá vlastní detekce

odlehých trajektorií kombinací přístupů založených na vzdálenosti a na hustotě. Trajektorie je považována za odlehlou, pokud obsahuje alespoň určitý počet odlehlých sub-trajektorií. Sub-trajektorie je odlehlá, pokud je blízká jen s určitým maximálním počtem trajektorií. Množina trajektorií, které jsou blízké ke konkrétní sub-trajektorii, je určena pomocí vzdálenostní funkce, která se skládá ze tří složek: kolmé, paralelní a úhlové vzdálenosti. Za účelem omezení vlivu hustoty oblasti na posuzování odlehlosti sub-trajektorie je zaveden tzv. upravující koeficient, kterým je násoben zjištěný počet blízkých trajektorií.

Odlisný přístup k detekci anomálií v datech pohybujících se objektů používá rámec nazvaný **ROAM**, který je navržen v [3]. Hlavní rozdíl spočívá ve využití principu klasifikace na základě trénovací množiny dat, ve které je každá trajektorie zařazena do odpovídající třídy. Nevýhodou tohoto přístupu je závislost na kvalitě trénovací datové množiny. Celý proces prochází postupně třemi moduly. Nejprve jsou ze vstupních trajektorií kombinací metody klouzavého okna a shlukování extrahovány společné vzory, tzv. motify. Každý motif reprezentuje nějaký typický vzor pohybu, například odbočení doprava, otočení a podobně. Každá trajektorie je pak reprezentována jako sekvence výskytů různých motifů. Výskyt motifu je určen množinou odpovídajících atributů, kterými jsou pozice a čas výskytu, případně některé další jako rychlost pohybu a podobně. Množina všech motifů tvoří hierarchický prostor rysů. Trajektorie se tak stává vektorem rysů v tomto novém prostoru. Takto transformované trajektorie jsou vstupem klasifikátoru založeného na pravidlech.

## 2.1 ALGORITMUS TOP-EYE

Pro detekci odlehlých trajektorií již v průběhu jejich vývoje je v [1] navržena metoda nazvaná **TOP-EYE**. Tato metoda umožňuje, při vhodně nastavených vstupních parametrech, identifikovat odlehlé trajektorie v reálném čase a vyvarovat se velkého množství planých poplachů, které mohou být způsobené například šumem ve vstupních datech.

Algoritmus uvažuje sledovaný prostor jako pravidelnou mřížku malých buněk. Každá buňka je dále členěna do 8 směrových výsečí, kde každá výseč odpovídá úhlu  $\pi/4$ . Cílem tohoto dělení je sumarizovat směrovou informaci trajektorií procházejících konkrétní buňkou jako vektor osmi hodnot, které vyjadřují pravděpodobnosti pohybu objektu v osmi různých směrech. Buňka  $g$  je pak reprezentována směrovým vektorem  $g = (p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8)$ , kde  $p_i$  je frekvence trajektorií, které se pohybovaly skrz buňku  $g$  ve směru výseče  $i$ .

Tato metoda tedy vyžaduje trénovací datovou množinu, na jejímž základě je vypočtena hustota jednotlivých buněk. Analýzou každé trajektorie z této datové množiny jsou navýšeny odpovídající hodnoty směrového vektoru u buněk, kterými daná trajektorie prochází. Za účelem dosažení kvalitních výsledků je vhodné, aby byla trénovací množina zbavena odlehlých trajektorií.

Pro každou novou trajektorii pak algoritmus průběžně počítá *okamžitou míru odlehlosti*. Míru odlehlosti je možné sledovat podle směru trajektorie nebo podle hustoty buněk. Novou trajektorii můžeme reprezentovat směrovým vektorem, například  $(q_1, \dots, q_k)$ , kde  $k$  je počet směrů pohybu uvnitř sledované buňky a každé  $q_j = 1/k$ . Míra odlehlosti podle směru je pak dána vzdáleností mezi sumarizovaným směrovým vektorem sledované buňky a směrovým vektorem nové trajektorie:

$$OScoreDir = 1 - \sum_{j=1}^k q_j \sum_{i=1}^8 p_i \cdot \cos \angle(v_j, v_i) \quad (1)$$

kde  $\cos \angle(v_j, v_i)$  je kosinus úhlu mezi směrem pohybu  $j$  nové trajektorie a reprezentativním směrem výseče  $i$ . Míra odlehlosti podle hustoty:

$$OScoreDen = \begin{cases} s & \text{jestliže } density < \tau \\ 0 & \text{jinak} \end{cases} \quad (2)$$

kde *density* je sumarizovaná hustota sledované buňky,  $s$  a  $\tau$  jsou uživatelské parametry.

Hlavní myšlenkou algoritmu TOP-EYE je postupně v čase akumulovat okamžité míry odlehlosti s tím, že vliv předchozích měř odlehlosti je snižován prostřednictvím exponenciální funkce  $\exp(-\lambda\Delta t)$  v závislosti na čase. Uvažujme, že  $S_{t_i}$  je okamžitá míra odlehlosti získaná pomocí rovnice (1) nebo rovnice (2). *Vyvíjející se míra odlehlosti* v časovém okamžiku  $t_i$  je pak dána jako

$$S_{t_i}^{\Sigma} = S_{t_i} + S_{t_{i-1}} \cdot \exp(-\lambda\Delta t_{i-1}) + S_{t_{i-2}} \cdot \exp(-\lambda\Delta t_{i-2}) + \dots + S_{t_0} \cdot \exp(-\lambda\Delta t_0) \quad (3)$$

kde  $\Delta t_{i-k}$  je časový rozdíl mezi časy  $t_i$  a  $t_{i-k}$ ,  $\lambda$  je uživatelský parametr.

Tímto způsobem je možné pozorovat postupný vývoj míry odlehlosti trajektorie v průběhu pohybu objektu. Při vhodně nastaveném parametru, který určuje hranici mezi odlehlými a normálními trajektoriemi, je tato míra odlehlosti odolná vůči planým poplachům.

### 3 APLIKACE PRO DETEKCI ODLEHLÝCH TRAJEKTORIÍ

V rámci tohoto projektu bude v programovacím jazyce Java implementován výše popsaný algoritmus TOP-EYE. Za účelem otestování, ověření a možnosti použití algoritmu pro detekci odlehlých trajektorií bude navržena a vytvořena vhodná aplikace s grafickým uživatelským rozhraním. Experimentální ověření a zhodnocení výsledné realizace algoritmu bude provedeno na reálné datové sadě trajektorií získané z multi-kamerového dohledového systému v rámci vývoje systému SUNAR [4] na FIT VUT v Brně.

### 4 ZÁVĚR

Tento článek se zabývá dolováním odlehlých trajektorií. Tato oblast má význam pro detekci a identifikaci anomálií v datech pohybujících se objektů, tedy neobvyklého či podezřelého chování. Článek představuje zástupce některých existujících přístupů se zaměřením na rozbor a implementaci algoritmu TOP-EYE, který poskytuje možnost účinně detekovat odlehlé trajektorie již v rané fázi jejich vývoje s ohledem na požadavek na nízkou míru planých poplachů.

### PODĚKOVÁNÍ

Tato práce vznikla částečně za podpory výzkumného záměru MSM0021630528 Výzkum informačních technologií z hlediska bezpečnosti.

### REFERENCE

- [1] Ge, Y., Xiong, H., Zhou, Z.-H., Ozdemir, H., Yu, J., Lee, K. C.: TOP-EYE: Top-k Evolving Trajectory Outlier Detection. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 2010, s. 1733–1736.
- [2] Lee, J.-G., Han, J., Li, X.: Trajectory Outlier Detection: A Partition-and-Detect Framework. In *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering*, 2008, s. 140–149.
- [3] Li, X., Han, J., Kim, S., Gonzalez, H.: ROAM: Rule- and Motif-Based Anomaly Detection in Massive Moving Object Data Sets. In *Proceedings of the 7th SIAM International Conference on Data Mining*, 2007.
- [4] Chmelař, P., Láník, A., Mlích, J.: SUNAR: Surveillance Network Augmented by Retrieval. In *ACIVS 2010*, s. 155–166.