



دانشگاه صنعتی شریف
دانشکده مهندسی کامپیوتر

رساله دکتری
هوش مصنوعی

عنوان رساله

تحلیل تاریخچه رویداد در شبکه‌های پیچیده

نگارش
علی زارع زاده

استاد راهنما
دکتر حمیدرضا ربیعی

تابستان ۱۳۹۷

دانشگاه صنعتی شریف
دانشکده مهندسی کامپیوتر

رساله دکتری

تحلیل تاریخچه رویداد در شبکه‌های پیچیده

نگارش: علی زارع زاده

استاد راهنما: دکتر حمیدرضا ربیعی امضاء:

داوران داخلی: دکتر جعفر حبیبی امضاء:

دکتر مهدیه سلیمانی امضاء:

داوران خارجی: دکتر مرتضی آنالویی امضاء:

دکتر مسعود اسدپور امضاء:

دکتر عمادالدین فاطمی زاده امضاء:

چکیده

تحلیل تاریخیچه رویدادها از مسایل کلاسیک در حوزه فرآیندهای تصادفی است که اخیراً در بستر شبکه‌های پیچیده مورد توجه زیادی قرار گرفته است. هدف از این مساله مدل کردن زمان وقوع یک رویداد، مانند رفتن به رستوران یا ارسال یک خبر، با در اختیار داشتن تاریخیچه آن و در مرحله بعد، کنترل و هدایت آن رویداد است. در رساله حاضر، تلاش کردیم در این دو بخش پیشرفت‌هایی حاصل کنیم...

کلمات کلیدی: شبکه‌های پیچیده، فرآیندهای تصادفی نقطه‌ای، فرآیند هاوکس، رویدادهای مکانی-زمانی، کنترل بهینه، معادلات دیفرانسیل تصادفی

فهرست مطالب

۲	فصل ۱	مقدمه
۲	۱-۱	تحلیل تاریخچه رویداد
۳	۲-۱	هدف پژوهش
۳	۳-۱	نوآوری های رساله
۳	۴-۱	ساختار رساله
۴	فصل ۲	پژوهش های پیشین
۴	۱-۲	مدل سازی انتشار رفتار
۴	۱-۱-۲	زمان پیوسته و بدون ساختار شبکه
۶	فصل ۳	مدل سازی رویداد
۶	۱-۳	انتشار رفتارهای همبسته
۶	۱-۱-۳	استنتاج پارامترها
۸	فصل ۴	آزمایش ها
۹	فصل ۵	جمع بندی
۱۰	پیوست آ	فرآیندهای نقطه ای
۱۱	۱-آ	تعریف فرآیند پواسن
۱۲	پیوست ب	اثبات ها

۱۴

مراجع

۱۶

واژه‌نامه فارسی به انگلیسی

۱۷

واژه‌نامه انگلیسی به فارسی

فهرست تصاویر

۱-۲	مدل‌های سرایت SI، SIS و SIR به ترتیب بالا سمت راست، بالا سمت چپ و پایین [۱].	۵
۱-آ	فرآیند پواسن یک‌بعدی	۱۰
۲-آ	فرآیند پواسن چندبعدی، استقلال آماری در توزیع نقاط	۱۱

فهرست جداول

۱-۳	پارامترهای مدل زمانی داده‌های ورود کاربران.	۷
-----	---	---

فصل ۱

مقدمه

تحلیل تاریخچه رویداد از مسایل کلاسیک آمار است که به مدل سازی، پیش بینی و (اخیرا) کنترل رویدادها می پردازد. منظور از رویداد^۲، یک اتفاق خاص در یک زمان مشخص است؛ مانند بیماری، زلزله، مشاهده یک خبر یا رفتن به یک مکان. هر رویداد با زمان وقوع (و نوع) آن مشخص می شود. رویدادها می توانند از نوع بقا^۳ یا بازگشتی^۴ باشند...

۱-۱ تحلیل تاریخچه رویداد

برای مدل سازی رویدادها (پدیده هایی که به صورت اتفاق هایی در زمان پیوسته رخ می دهند) از فرآیندهای تصادفی نقطه ای^۵ استفاده می شود. فرآیندهای نقطه ای، برای مدل سازی در گستره وسیعی از کاربردهای شبکه های اجتماعی و سیستم های اطلاعاتی مانند؛ انتشار اطلاعات [۲، ۳، ۴]، دینامک نظرات [۵]، رقابت محصولات [۶، ۷]، اتکاپذیری اطلاعات [۸] یا یادگیری انسان ها [۹] استفاده شده است.

یک فرآیند تصادفی را می توان با دانستن توزیع توام متناهی اش به طور کامل بیان کرد (پیوست آ، قضیه کولموگروف^۶). توزیع توام متناهی $f(t_1, t_2, \dots, t_n)$ ، چگالی احتمال وقوع رویدادها در زمان های t_1 تا t_n دلخواه را بیان می کند. طبق قانون بیز می توان این توزیع توام را به فرم زیر نوشت:

$$f(t_1, t_2, \dots, t_n) = \prod_i f(t_i | t_{i-1}) \quad (1.1)$$

² Event

³ Survival

⁴ Recurrent

⁵ Stochastic point process

⁶ Kolmogorov

۲-۱ هدف پژوهش

هدف از این پژوهش، مدل‌سازی و کنترل رویدادها با در نظر گرفتن فرضیات مساله و وجود ساختار شبکه است. در واقع هر یک از گره‌های شبکه مجموعه‌ای از رویدادها تولید می‌کند، و در مساله مدل‌سازی به دنبال این هستیم که از فرضیات مساله و تعامل بین گره‌ها و تاریخیچه رویدادهایشان برای تعیین رویداد بعدی هر کاربر در شبکه استفاده کنیم...

۳-۱ نوآوری‌های رساله

نوآوری‌های این رساله در بخش مدل‌سازی و کنترل رویداد شامل...

۴-۱ ساختار رساله

در ادامه این رساله ابتدا در فصل ۲، به پژوهش‌های پیشین در زمینه مدل‌سازی و کنترل رویدادها می‌پردازیم. در فصل ...

فصل ۲

پژوهش‌های پیشین

در این فصل پژوهش‌های مرتبط با ...

۱-۲ مدل‌سازی انتشار رفتار

رفتارهای اجتماعی، انتشار نرُم‌های فرهنگی و حتی اجماع بین افراد معمولاً به صورت واکنش‌های دینامیکی بین مجموعه‌ای از عامل‌های به هم متصل رخ می‌دهد [۱۰]. مدل‌سازی این دینامیک روی شبکه توجه محققان زیادی را در چند سال اخیر به خود جلب کرده است. با توجه به ساختار شبکه و پیوستگی زمان، این مدل‌ها را می‌توان به سه دسته تقسیم کرد؛ زمان پیوسته و بدون ساختار شبکه، زمان گسسته و با ساختار شبکه، و زمان پیوسته و با ساختار شبکه. در ادامه پژوهش‌های انجام شده در هر سه دسته آورده می‌شود.

۱-۱-۲ زمان پیوسته و بدون ساختار شبکه

این مدل‌ها در ابتدا برای تشخیص سرایت بیماری‌ها ارائه شده‌اند. چارچوب مدل‌های ارائه شده بر دو فرض اصلی استوار است؛ حالت چندبخشی^۲ افراد و ترکیب همگن^۳ جامعه [۱]. در مدل‌های سرایت افراد بر اساس مرحله پیشروی بیماری به حالت‌های مختلفی، مانند موارد زیر، تقسیم می‌شوند.

- مستعد^۴ (S): فرد سالمی که هنوز در معرض عامل بیماری‌زا قرار نگرفته است.

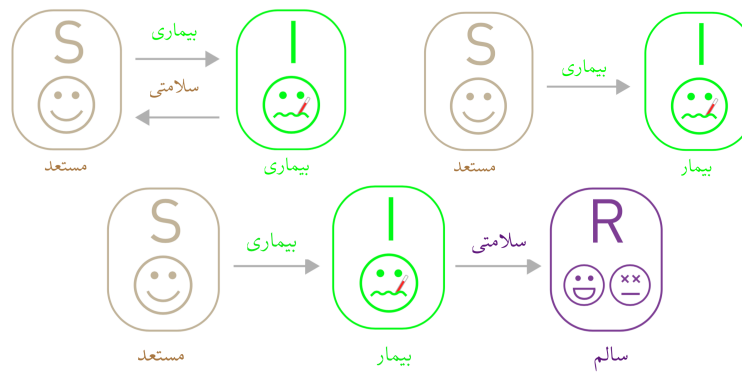
- بیمار^۵ (I): فرد مبتلا به بیماری که می‌تواند بیماری را انتقال دهد.

² Compartmentalization

³ Homogenous Mixing

⁴ Susceptible

⁵ Infectious



شکل ۲-۱: مدل‌های سرایت SI، SIS و SIR به ترتیب بالا سمت راست، بالا سمت چپ و پایین [۱].

- سالم^۱ (R): فردی که قبلاً بیمار شده است و اکنون ناقل بیماری نیست.

ترکیب همگن افراد نیز بدین صورت است که گراف ساختار ندارد و هر فرد با احتمال یکسانی به افراد بیمار اتصال دارد. در واقع افراد به صورت همگن در شبکه به یکدیگر متصل هستند. ساده‌ترین مدل سرایت، SI نام دارد که...

¹ Recovered

فصل ۳

مدل سازی رویداد

در پژوهش های اولیه ی مدل سازی و تحلیل تاریخچه رویداد، مفهومی به نام شبکه کمتر به چشم می خورد...

۱-۳ انتشار رفتارهای همبسته

به رد گذاشته شده از انتشار رفتار، اطلاعات، ویروس، بیماری، اخبار یا علاقه مندی کاربران به یک محتوا در شبکه ها، انتشار (آبشار) می گوئیم [۱۱]...

۱-۱-۳ استنتاج پارامترها

برای به دست آوردن درستنمایی مدل، از گزاره زیر استفاده می کنیم، که به طور کلی می توان از آن برای به دست آوردن درستنمایی فرایندهای نقطه ای نشان دار استفاده کرد.

گزاره ۱. فرآیند نقطه ای نشان دار چند متغیره $N, 1, 2, \dots, u$ را با تابع شدت $\lambda_u(t)$ و نشان $f_u(p|t)$ در نظر بگیرید. فرض کنید $\mathcal{D} = \{(t_i, u_i, p_i)\}_{i=1}^K$ یک وقوع از این فرآیند تصادفی در بازه $[0, T]$ باشد. آنگاه درستنمایی \mathcal{D} مدل N_u با پارامترهای θ را می توان به صورت زیر محاسبه کرد:

$$f(\mathcal{D}|\theta) = \left[\prod_{i=1}^K \lambda_{u_i}(t_i) f_{u_i}(p_i|t_i) \right] \exp \left(- \int_0^T \sum_{u=1}^N \lambda_u(s) ds \right)$$

اثبات. به پیوست ب، مراجعه کنید.

جدول ۳-۱: پارامترهای مدل زمانی داده‌های ورود کاربران.

پارامتر	توضیح
β_u	پارامتر هسته u
μ_{uc}	شدت پایه کاربر u در دسته c
α_{vu}	تأثیر کاربر u بر v
η_{uc}	تمایل کاربر u به اکتشاف مکان‌های جدید از دسته c

فصل ۴

آزمایش‌ها

در این بخش الگوریتم پیشنهادی را بر روی مجموعه داده‌های واقعی و ساختگی مورد ارزیابی قرار می‌دهیم...

فصل ۵

جمع بندی

در این رساله به تحلیل تاریخچه رویداد در شبکه‌های پیچیده در دو بخش مدل‌سازی و کنترل رویداد پرداختیم. ...

این روش راه را برای کارهای آتی مختلفی باز می‌کند. به عنوان مثال، از منظر یک کاربر، به شبکه‌های اجتماعی (مانند توییتر و ویبو) پرداختیم که فید کاربران را به صورت معکوس زمانی مرتب می‌کنند. تعمیم این روش به شبکه‌های دیگر مانند فیسبوک که الگوریتم خاصی برای این منظور دارد، یکی از کارهای آتی است. ...

پیوست آ

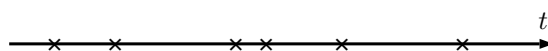
فرآیندهای نقطه‌ای

یکی از معروف‌ترین توزیع‌ها در آمار و احتمال، توزیع پواسن است که حالت حدی توزیع دو جمله‌ای است وقتی که تعداد آزمایش‌ها زیاد و احتمال موفقیت کم باشد. اگر تعداد متوسط موفقیت‌ها را $\mu = Np$ بنامیم می‌توان نشان داد:

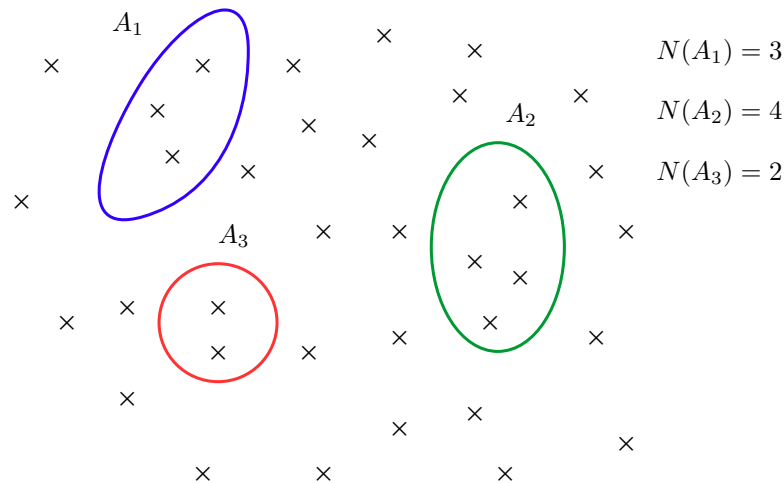
$$\text{Pois}(r|\mu) = \lim_{n \rightarrow \infty} \text{Bin}(r|N, p) = \frac{\mu^r e^{-\mu}}{r!} \quad (1.A)$$

که μ میانگین توزیع پواسن نیز است. به طور مشابه فرآیند پواسن برای شمارش پدیده‌هایی مانند تابش ذرات رادیواکتیو، تماس‌های گرفته شده با مرکز تلفن یا درخواست‌ها از یک وب‌سرور کار می‌رود که به صورت رویدادهایی مستقل در زمان پیوسته اتفاق می‌افتند، شکل آ-۱ را ببینید. در حالت چندبُعدی می‌توان توزیع ستارگان در آسمان یا درختان در جنگل را که هیچ الگوی نظم خاصی ندارد مانند شکل آ-۲ با فرآیند پواسن مدل کرد. در واقع پدیده‌هایی که از عوامل مستقل زیادی به وجود می‌آیند که هر کدام احتمال کمی دارند، به خوبی با فرآیند پواسن مدل می‌شوند. ویژگی اصلی این فرآیند تصادفی استقلال آماری آن است به طوری که تعداد نقاط در ناحیه‌هایی که با هم اشتراک ندارند از هم مستقل هستند.

در این بخش ابتدا تعریف و خواص توزیع پواسن آورده می‌شود. سپس قضایای مهم در مورد فرآیند پواسن بیان می‌شود. در بخش بعد انواع فرآیندهایی که از روی پواسن تعریف می‌شوند مانند فرآیند پواسن نشان‌دار، فرآیند هاوکس و فرآیند کاوکس آورده می‌شود. در انتها دواروش نمونه برداری اوگاتا و باریک‌سازی شرح داده می‌شود.



شکل آ-۱: فرآیند پواسن یک‌بُعدی



شکل آ-۲: فرآیند پواسن چندبُعدی، استقلال آماری در توزیع نقاط

۱-آ تعریف فرآیند پواسن

برای تعریف فرآیندهای تصادفی دو دیدگاه وجود دارد؛ مجموعه متغیرهای تصادفی و تابع تصادفی. برای تعریف فرآیند تصادفی ابتدا متغیر تصادفی را تعریف می‌کنیم [۱۲].

تعریف ۲. متغیر تصادفی X تابعی اندازه‌پذیر از فضای احتمال (Ω, \mathcal{F}, P) به فضای اندازه‌پذیر (Ξ, \mathcal{E}) است بدین معنا که نگاشت معکوس $E \in \mathcal{E}$ عضو \mathcal{F} است، $X^{-1}(E) \in \mathcal{F}$. برای تعریف توزیع احتمال متغیر تصادفی، فضای اندازه‌پذیر را $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ در نظر می‌گیرند^b. اکنون توزیع تجمعی را می‌توان به صورت $F_X(x) = P(X^{-1}(-\infty, x]) = P(\{\omega | X(\omega) \leq x\})$ نوشت.

^a Measurable Space

^b مجموعه $\mathcal{B}(\mathbb{R})$ از کامل کردن $\{(-\infty, q) | q \in \mathbb{Q}\}$ به دست می‌آید، یعنی کوچکترین میدان سیگمایی که مجموعه نیم‌بازه‌های کسری عضو آن باشند.

از اینجا به بعد فرض می‌کنیم فضای احتمال (Ω, \mathcal{F}, P) را در اختیار داریم که همه متغیرهای تصادفی در آن قابل تعریف هستند. اکنون تعریف فرآیند تصادفی به صورت مجموعه‌ای از متغیرهای تصادفی را می‌توان بیان کرد [۱۳].

تعریف ۳. فرآیند تصادفی $\{X_t\}_{t \in \mathcal{T}}$ مجموعه‌ای از متغیرهای تصادفی X_t از فضای احتمال (Ω, \mathcal{F}, P) به فضای اندازه‌پذیر (Ξ, \mathcal{E}) است که با مجموعه \mathcal{T} نمایه می‌شوند.

برای بیان تعریف دوم، باید ابتدا تابع تصادفی و نمونه مسیر^۱ را تعریف کنیم [۱۳].

¹ Sample path

پیوست ب

اثبات‌ها

اثبات گزاره ۱

با استفاده از قانون زنجیر در احتمالات می‌توان نوشت

$$f(\mathcal{D}|\theta) = \prod_{i=1}^K f((t_i, u_i, p_i)|\mathcal{D}(t_i)) \prod_{u=1}^N S(T, u)$$

که $t_0 = 0$ و $S_u(T)$ احتمال بقای فرآیند $\lambda_u(t)$ بعد از آخرین رویدادش است.

$$S_u(T) = \exp\left(-\int_{t_{|\mathcal{D}_u|}}^T \lambda_u(s)ds\right)$$

اکنون با استفاده از رابطه فوق می‌توان درست‌نمایی را محاسبه کرد.

$$\begin{aligned} f(\mathcal{D}|\theta) &= \prod_{u=1}^N \prod_{i=1}^{|\mathcal{D}_u|} f((t_i, u_i, p_i)|\mathcal{D}(t_i)) \prod_{u=1}^N S(T, u) \\ &= \prod_{u=1}^N \prod_{i=1}^{|\mathcal{D}_u|} \lambda_u(t_i) \exp\left(-\int_{t_{i-1}}^{t_i} \lambda_u(s)ds\right) f_u(p_i|t_i) \prod_{u=1}^N S(T, u) \end{aligned}$$

$$\begin{aligned}
 &= \prod_{u=1}^N \exp \left(- \int_0^{t_{|\mathcal{D}_u|}} \lambda_u(s) ds \right) \prod_{i=1}^{|\mathcal{D}_u|} f_u(p_i|t_i) \lambda_u(t_i) \prod_{u=1}^N S(T, u) \\
 &= \prod_{u=1}^N \exp \left(- \int_0^{t_{|\mathcal{D}_u|}} \lambda_u(s) ds \right) S(T, u) \prod_{i=1}^{|\mathcal{D}_u|} f_u(p_i|t_i) \lambda_u(t_i) \\
 &= \prod_{u=1}^N \exp \left(- \int_0^T \lambda_u(s) ds \right) \prod_{i=1}^{|\mathcal{D}_u|} f_u(p_i|t_i) \lambda_u(t_i) \\
 &= \prod_{u=1}^N \exp \left(- \int_0^T \lambda_u(s) ds \right) \prod_{u=1}^N \prod_{i=1}^{|\mathcal{D}_u|} f_u(p_i|t_i) \lambda_u(t_i) \\
 &= \exp \left(- \int_0^T \sum_{u=1}^N \lambda_u(s) ds \right) \prod_{i=1}^K \lambda_{u_i}(t_i) f_{u_i}(p_i|t_i)
 \end{aligned}$$

مراجع

- [1] A. L. Barabasi, *Network Science*. Cambridge university press, 2015.
- [2] M. G. Rodriguez, D. Balduzzi, and B. Schölkopf, “Uncovering the temporal dynamics of diffusion networks,” in *Proceedings of the 28th International Conference on Machine Learning*, 2011, pp. 561–568.
- [3] N. Du, L. Song, M. Gomez-Rodriguez, and H. Zha, “Scalable influence estimation in continuous-time diffusion networks,” in *Advances in Neural Information Processing Systems*, 2013, pp. 3147–3155.
- [4] Q. Zhao, M. Erdogdu, H. He, A. Rajaraman, and J. Leskovec, “Seismic: A self-exciting point process model for predicting tweet popularity,” in *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 1513–1522.
- [5] A. De, I. Valera, N. Ganguly, S. Bhattacharya, and M. Gomez-Rodriguez, “Learning and forecasting opinion dynamics in social networks,” in *Advances in Neural Information Processing Systems*, 2016, pp. 397–405.
- [6] I. Valera and M. Gomez-Rodriguez, “Modeling diffusion of competing products and conventions in social media,” in *IEEE International Conference on Data Mining*, 2015, pp. 409–418.
- [7] A. Zarezade, A. Khodadadi, M. Farajtabar, H. R. Rabiee, and H. Zha, “Correlated cascades: Compete or cooperate,” in *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, 2017, pp. 238–244.
- [8] B. Tabibian, I. Valera, M. Farajtabar, L. Song, B. Schoelkopf, and M. Gomez-Rodriguez, “Distilling information reliability and source trustworthiness from digital traces,” in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 847–855.

- [9] C. Mavroforakis, I. Valera, and M. Gomez-Rodriguez, “Modeling the dynamics of online learning activity,” in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 1421–1430.
- [10] A. Vespignani, “Modelling dynamical processes in complex socio-technical systems,” *Nature Physics*, vol. 8, no. 1, pp. 32–39, 2012.
- [11] M. Gomez-Rodriguez, L. Song, H. Daneshmand, and B. Schölkopf, “Estimating diffusion networks: Recovery conditions, sample complexity and soft-thresholding algorithm,” *Journal of Machine Learning Research*, vol. 17, pp. 90:1–90:29, 2016.
- [12] D. Williams, *Probability with martingales*. Cambridge university press, 1991.
- [13] C. Shalizi and A. Kontorovich, “Almost none of the theory of stochastic processes a course on random processes, 2010.”

واژه‌نامه فارسی به انگلیسی

Stochastic point process	فرآیندهای تصادفی نقطه‌ای .	Recurrent	بازگشتی
Measurable Space	فضای اندازه‌پذیر	Survival	بقا
Susceptible	مستعد	Infectious	بیمار
Sample path	نمونه مسیر	Homogenous Mixing	ترکیب همگن
Compartmentalization	چندبخشی	Event	رویداد
		Recovered	سالم

واژه‌نامه انگلیسی به فارسی

Recurrent	بازگشتی	Compartmentalization	چندبخشی
Sample path	نمونه مسیر	Event	رویداد
Stochastic point process	فرآیندهای تصادفی	Homogenous Mixing	ترکیب همگن
	نقطه‌ای	Infectious	بیمار
Survival	بقا	Measurable Space	فضای اندازه‌پذیر
Susceptible	مستعد	Recovered	سالم

Abstract

Event history analysis is a classic problem in stochastic process, that recently attracts the researchers' attention in complex networks. The goal is to model the time of events, like check in a restaurant check-in or post a message, given their history, and then control them. In this thesis we try to improve both aspects of the problem...

Keywords: *Event history, complex network, stochastic point process, Hawkes process, spatio-temporal events, optimal control, stochastic differential equations*



Sharif University of Technology
Department of Computer Engineering

Ph.D. Thesis
Artificial Intelligence

Topic
Event History Analysis in Complex Networks

By
Ali Zarezade

Supervisor
Prof. Hamid R. Rabiee

Summer 2018