ریاضی و آمار

مدل‌های احتمال برای داده‌های شمارشی

نوشتهٔ مهرداد کیانی — دانشیار آمار و مدل‌سازی آماری، دانشکدهٔ آمار و انفورماتیک، دانشگاه شهر نیلگون

در بسیاری از آزمایش‌ها داده‌ها به صورت شمارش ظاهر می‌شوند: تعداد تصادف‌ها در یک چهارراه در یک ساعت، تعداد مشتری‌های یک فروشگاه در روز، یا تعداد اشتباه‌های چاپی در یک صفحه. در این حالت‌ها متغیر تصادفی گسسته است و مقدارهای ممکن آن اعداد صحیح نامنفی‌اند. در این مقاله سه مدل پرکاربرد برای چنین داده‌هایی معرفی می‌شود و شرط به‌کارگیری هر یک با مثال توضیح داده می‌شود.

مدل پواسن برای شمارش رویدادهای نادر

اگر رویدادها به طور مستقل و با نرخ ثابت در زمان رخ دهند، تعداد رویدادها در یک بازهٔ زمانی ثابت از توزیع پواسن با پارامتر λ پیروی می‌کند. شرط λ برابر میانگین و واریانس است. در عمل λ را از داده‌های گذشته برآورد می‌کنیم و سپس با آزمون نیکویی برازش بررسی می‌کنیم آیا داده‌ها با مدل سازگارند یا خیر.

مدل دوجمله‌ای منفی برای داده‌های بیش‌پراکنده

وقتی واریانس داده‌ها بزرگ‌تر از میانگین است، مدل پواسن نامناسب است. در این حالت مدل دوجمله‌ای منفی با دو پارامتر r و p داده‌ها را بهتر توصیف می‌کند. r می‌تواند حتی عدد صحیح نباشد و p احتمال موفقیت در هر آزمایش است. برآورد پارامترها با روش گشتاورها ساده است و برای مقایسه با مدل پواسن نیز به کار می‌رود.

مدل دوجمله‌ای برای نسبت‌ها

اگر تعداد آزمایش‌ها n باشد و احتمال موفقیت p ثابت بماند، تعداد موفقیت‌ها X از توزیع دوجمله‌ای پیروی می‌کند. میانگین np و واریانس $np(1-p)$ است. وقتی n بزرگ و p کوچک است، توزیع پواسن با $\lambda = np$ تقریب خوبی می‌دهد. این تقریب برای محاسبه‌های دستی بسیار کارآمد است.

انتخاب مدل از روی نمودار

نمودار فراوانی داده‌ها و نمودار نسبت به مقدار مشاهده‌شده کمک می‌کند مدل مناسب انتخاب شود. اگر نمودار به سمت راست کشیده باشد و دم بلند داشته باشد، مدل دوجمله‌ای منفی محتمل‌تر است. اگر نمودار متقارن و تیز باشد، پواسن یا دوجمله‌ای مناسب‌تر است. این انتخاب باید همراه با آزمون نیکویی برازش تأیید شود.

شبکه‌های اجتماعی

شبکه‌های اجتماعی و انتشار شایعه

نوشتهٔ مهرداد کیانی

وقتی خبری در یک شبکه پخش می‌شود، همه‌ی کاربران به یک اندازه آن را باور نمی‌کنند و همه به یک اندازه آن را پخش نمی‌کنند. این مقاله نشان می‌دهد چرا برخی شایعه‌ها گسترده می‌شوند و چرا برخی دیگر در همان ابتدا می‌میرند. ما از مدل ساده‌ای استفاده می‌کنیم که هم برای آموزش مناسب است و هم برای تحلیل ریاضی.

مدل آبشاری مستقل

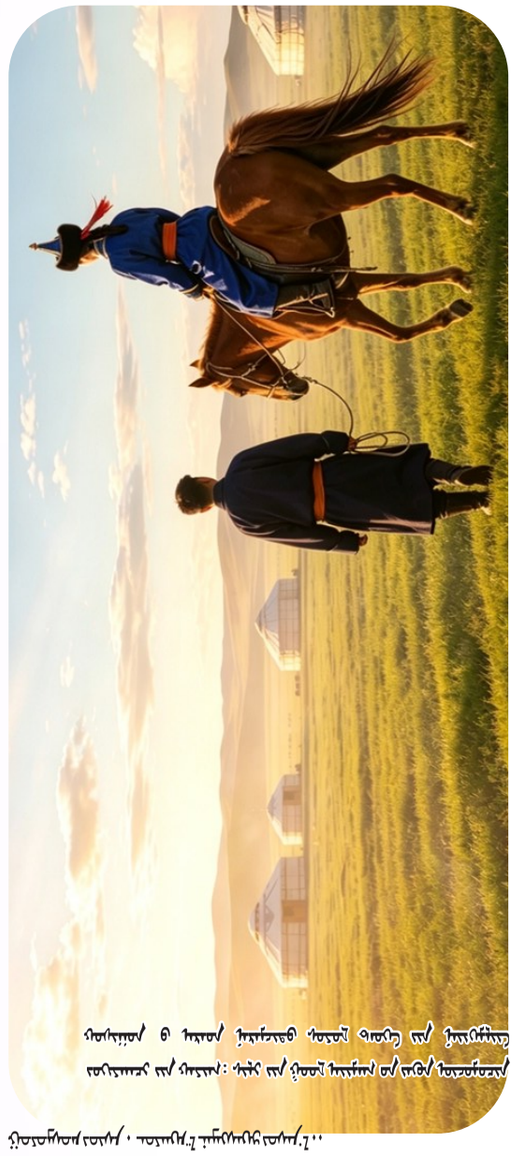
در مدل آبشاری مستقل، هر کاربر یک بار تصمیم می‌گیرد شایعه را بپذیرد یا نپذیرد. اگر بپذیرد، با احتمال ثابت p آن را به همسایه‌هایش می‌فرستد و سپس از شبکه خارج می‌شود. این مدل ساده است و امکان می‌دهد اندازه‌ی نهایی شایعه را برای شبکه‌های تصادفی حساب کنیم. نقطه‌ی بحرانی $p_c = 1/\langle k \rangle$ است که در آن اندازه‌ی شایعه از کوچک به بزرگ تغییر می‌کند.

مدل آستانه‌ای

در مدل آستانه‌ای، کاربر تنها زمانی شایعه را می‌پذیرد که تعداد همسایه‌های پذیرنده‌ی او از آستانه‌ی θ بیشتر شود. این مدل برای شایعه‌هایی مناسب است که به اعتماد چند منبع نیاز دارند، مانند تصمیم‌های خرید یا اعتراض‌های اجتماعی. ما نشان می‌دهیم که با $\theta = 2$ و شبکه‌ی تصادفی، شایعه می‌تواند ناگهان گسترده شود.

شبکه‌های واقعی

داده‌های واقعی نشان می‌دهند که شبکه‌های اجتماعی دارای خوشه‌بندی بالا و طول مسیر کوتاه‌اند. این ویژگی‌ها هر دو مدل بالا را تغییر می‌دهند. ما نتایج شبیه‌سازی را برای شبکه‌ی تصادفی و شبکه‌ی کوچک‌جهان مقایسه می‌کنیم و نشان می‌دهیم که خوشه‌بندی چگونه سرعت پخش را کاهش می‌دهد.



دانشجویان در آزمایشگاه داده‌های اجتماعی دانشگاه شهر نیلگون هنگام تحلیل شبکه‌ی دوستی کلاس

گزارش ویژه: بازار مسکن

قیمت مسکن در شهر نیلگون ۱۴۰۳–۱۴۰۴

نوشتهٔ مهرداد کیانی

بازار مسکن شهر نیلگون در دو سال گذشته مسیری پرنوسان را پشت سر گذاشت. در بهار ۱۴۰۳ قیمت هر متر مربع آپارتمان در مناطق مرکزی به حدود ۴۲ میلیون تومان رسید و در پاییز همان سال به ۳۸ میلیون کاهش یافت. این گزارش روند قیمت‌ها، حجم معاملات و عوامل اصلی تغییر را بر پایه داده‌های دفاتر املاک و آمار بانک مرکزی بررسی می‌کند.

روند قیمت‌ها

میانگین قیمت هر متر مربع در مناطق مرکزی و شمالی شهر در این دوره ۱۲ درصد افزایش یافت، اما در مناطق حاشیه‌ای ۹ درصد کاهش داشت. شاخص کل بازار در پایان دی ۱۴۰۴ برابر ۱۷۴ بود که نسبت به پایه ۱۰۰ در ابتدای ۱۴۰۳ رشد معنادار نشان می‌دهد. نوسان ماهانه در نیمه دوم سال به ۶ درصد رسید که بیشتر از میانگین پنج سال گذشته است.

حجم معاملات

تعداد معاملات ثبت‌شده در این دوره ۱۸٬۴۰۰ فقره بود که نسبت به دوره‌ی مشابه ۲۰ درصد کمتر است. کاهش حجم معاملات در مناطق حاشیه‌ای بیشتر بود و نشان می‌دهد خریداران در انتظار ثبات قیمت‌اند. مدت زمان فروش یک واحد مسکونی از ۴۱ روز به ۶۷ روز افزایش یافت و این افزایش بیشتر بر واحدهای بزرگ‌متراژ اثر گذاشت.

عوامل اثرگذار

سه عامل اصلی تغییرات بازار عبارت‌اند از: نرخ سود تسهیلات، حجم نقدینگی و عرضه‌ی واحدهای نوساز. افزایش نرخ سود تسهیلات خرید را گران‌تر کرد و تقاضای واحدهای بزرگ را کاهش داد. همزمان تحویل حدود ۶٬۲۰۰ واحد نوساز در شهرک‌های اطراف فشار عرضه را افزایش داد و بر قیمت‌ها در حاشیه اثر گذاشت.

پیش‌بینی

با فرض ثبات نرخ سود و تحویل واحدهای در دست ساخت، پیش‌بینی می‌شود قیمت‌ها در نیمه‌ی نخست ۱۴۰۵ با نوسان محدود حول سطح فعلی حرکت کنند و تا پایان سال حداکثر ۵ درصد افزایش یابند. اگر نرخ سود کاهش یابد یا حجم نقدینگی افزایش یابد، این پیش‌بینی باید بازنگری شود. این گزارش بر پایه‌ی داده‌های موجود تهیه شده و به‌روزرسانی آن در تیر ۱۴۰۵ انجام می‌شود.

پرسش‌های آزمون پایان ترم

(پاسخ: ۵ صفحه بعد)

۱. در مدل پواسن با نرخ $\lambda = 4$ در ساعت، احتمال دریافت دقیقاً ۶ تماس در یک ساعت را محاسبه کنید و تفاوت آن با احتمال دریافت ۴ تماس را توضیح دهید.

۲. داده‌های زیر تعداد اشتباه‌های چاپی در ۱۰ صفحه‌ی یک کتاب‌اند: ۰، ۱، ۲، ۰، ۳، ۱، ۰، ۴، ۱، ۰. میانگین و واریانس را محاسبه کنید و بگویید آیا مدل پواسن مناسب است یا مدل دوجمله‌ای منفی.

۳. در مدل آبشاری مستقل با $p = 0.2$ و شبکه‌ی تصادفی با $\langle k \rangle = 4$، نقطه‌ی بحرانی p_c را محاسبه کنید و توضیح دهید چرا شایعه در $p < p_c$ کوچک می‌ماند.

۴. در مدل آستانه‌ای با $\theta = 2$ و شبکه‌ی کوچک‌جهان، چرا خوشه‌بندی بالا سرعت پخش را کاهش می‌دهد؟ پاسخ را با مثال توضیح دهید.

۵. در بازار مسکن شهر نیلگون، قیمت هر متر مربع از ۴۲ میلیون به ۳۸ میلیون کاهش یافت. درصد کاهش را محاسبه کنید و بگویید آیا این کاهش از نظر آماری بزرگ است یا خیر.

۶. حجم معاملات در مناطق حاشیه‌ای ۲۰ درصد کمتر است اما مدت فروش ۶۷ روز شده است. این دو عدد را با هم تفسیر کنید و بگویید کدام عامل احتمالاً غالب است.

۷. پیش‌بینی شده است که ۶٬۲۰۰ واحد نوساز در شهرک‌های اطراف تحویل داده شود. چرا این عرضه می‌تواند فشار بر قیمت‌ها را کاهش دهد؟ دو دلیل بنویسید.

۸. اگر نرخ سود تسهیلات ۲ درصد کاهش یابد و حجم نقدینگی ثابت بماند، انتظار دارید قیمت‌ها در نیمه‌ی نخست ۱۴۰۵ چه تغییری کنند؟ پاسخ را با استدلال بنویسید.