

پیش‌بینی گراف انتشار در شبکه اجتماعی تلگرام



دانشجو: صدف نیکنام
استاد راهنما: دکتر مسعود اسدپور
دانشکده مهندسی برق و کامپیوتر، دانشگاه تهران



مقدمه

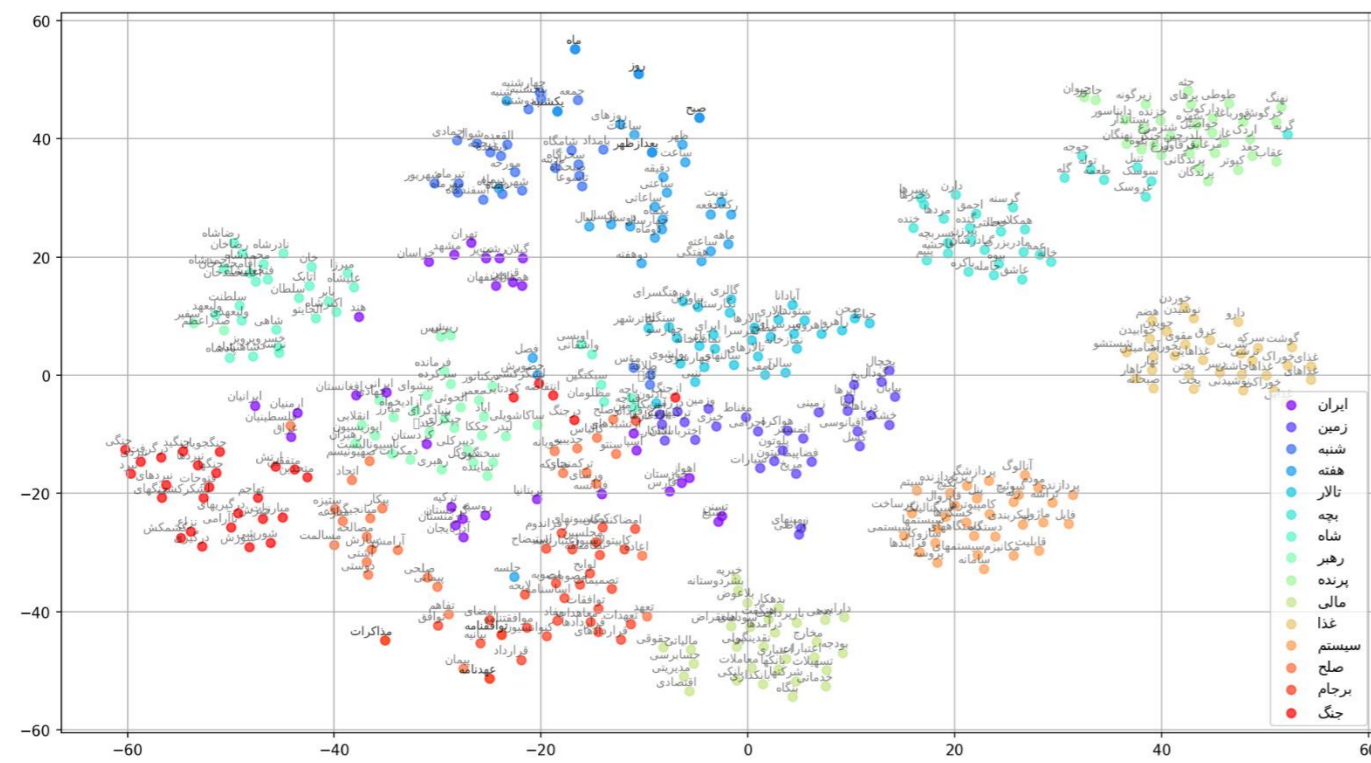
در عصر حاضر یکی از مهم‌ترین دستاوردهای انقلاب اطلاعاتی، شبکه‌های اجتماعی است که ارتباطات را معنایی تازه بخشیده و موجب تغییرات اساسی در تعاملات افراد شده است. ظهور این پدیده، هرچند معلول تحولاتی بوده اما خود منشا تهدیدات عدیده‌ای شده است.

تلگرام، یک شبکه اجتماعی معروف و پر کاربرد مخصوصا میان ایرانیان است که در آن، افراد عقاید و نقطه نظرات خود را به طرق مختلف بیان می‌کنند. با توجه به این مسئله، این شبکه اجتماعی می‌تواند منبع مناسبی برای پیدا کردن متون مختلف اجتماعی، سیاسی و فرهنگی باشد.

در این پژوهش با بررسی رفتار گذشته کاربران در مدت زمان مشخص و روابط میان آنها، سعی بر پیش‌بینی نحوه انتشار اخبار داریم.

نتایج

در شکل زیر نمایش برداری روی مجموعه داده ویکیپدیا برای تعدادی کلمه، دسته بندی و میزان شباهت آن‌ها به یکدیگر را میتوان مشاهده کرد.



برای متن تصادفی "آیت الله قاضی ره اگر کسی نماز واجبش را اول وقت بخواند به مقامات عالیہ نرسد مرا لعن کند کانال حکمت اسرار آل محمد" نتایج میزان تمایل کاربران به باز نشر پیام آورده شده است.

میزان تمایل به باز نشر پیام	متن ارسال شده توسط کاربر در گذشته	شناسه کاربر پیشنهادی
76%	آیت الله سید علی آقا قاضی کسی که نمازش را اول وقت بخواند وبه مقامات عالیہ نرسد مرا لعن کند کانال عاشقان ولیعصر عج	1186970
34%	امام کاظم علیه السلام از رسول الله صلی الله علیه آله نقل فرمودند هر کس دوست دارد خدا را ملاقات کند درحالی که خدا بسیار راحت به حسابش رسیدگی کند او را وارد بهشت جاویدانی کند که مساحتش اندازه آسمانها زمین است برای متقین آماده شده پس ولایت علی بن محمد الهادی را داشته باشد بحار الانوار	1097220
-58%	هر چه در خاطرم آید تو از آن خوبتری	761762

کاربرد

گراف یک شبکه اجتماعی برای مدل سازی ارتباط بین افراد استفاده می‌شود. با یافتن روابط میان افراد و پیام‌هایی که از یکدیگر باز نشر می‌کنند، میتوان رفتار آن‌ها را پیش‌بینی کرده و در نهایت از شیوع اخبار شایعه و مطالب نامناسب توسط افراد مشکوک در این شبکه‌ها جلوگیری کرد.

روش پیشنهادی

در این پژوهش نمایش برداری روی مجموعه داده کاربر و متون استخراج شده از شبکه تلگرام بررسی شد. مجموعه کاربران به صورت $u = \{u_1, u_2, \dots, u_n\}$ و مجموعه متون $d = \{d_1, d_2, \dots, d_n\}$ است. $d(u_i)$ مجموعه پیام‌هایی است که کانال u_i باز نشر می‌کند که از آن برای یادگیری شبکه مورد نظرم استفاده می‌شود. در کنار پیش‌بینی کلمات موجود در یک متن، لازم است تا میزان احتمال کاربر به متن را نیز بدست آورده شود و در نهایت هدف ما به حداکثر رساندن احتمال زیر است.

$$\frac{1}{T} \sum_t [\log p(w_t | d_t, w_{t-k}, \dots, w_{t+k}) + \log p(d_i | u_{i_1}, \dots, u_{i_n})]$$

جمع بندی

در این تحقیق روش‌های تعبیه متن و گراف در شبکه‌های اجتماعی را روی دو نوع مجموعه داده، بررسی کردیم و در نهایت برای کاربر u_k و متن d_j داده شده، شباهت کسینوسی میان نمایش برداری آن‌ها را محاسبه و پیش‌بینی میکنیم آیا کاربر تمایلی به باز نشر پیام دارد یا خیر.

با توجه به کوچک بودن مجموعه داده و همچنین کمبود داده برای زبان فارسی، لازم است تا تحقیقات را روی بازه زمانی بلندتر با موضوعات گسترده‌تر اجرا کرد تا در نهایت بتوان با دقت بالاتری نتیجه بهتری گرفت.

مراجع اصلی

- Grover, Aditya & Leskovec, Jure. (2016). node2vec: Scalable Feature Learning for Networks. KDD : proceedings. International Conference on Knowledge Discovery & Data Mining. 2016. 855-864. 10.1145/2939672.2939754.
- Yu, Yang & Wan, Xiaojun & Zhou, Xinjie. (2016). User Embedding for Scholarly Microblog Recommendation. 449-453. 10.18653/v1/P16-2073.
- Silvio Amir, Byron C Wallace, Hao Lyu, and Paula Carvalho Mario J Silva. 2016. Modelling context with user embeddings for sarcasm detection in social media. arXiv preprint arXiv:1607.00976.

