موضوع:

Detecting Deceptive Reviews using Generative Adversarial Networks

تهیه کننده:

میلاد دشتبانی

نام استاد:

دکتر زارع

شماره دانشجویی:

140112459009

سال تحصیلی:

1402

در چند سال گذشته، سایت‌های بررسی مصرف‌کننده به هدف اصلی هرزنامه‌های فریبنده نظرات تبدیل شده‌اند، جایی که نظرات یا بررسی‌های ساختگی عمداً نوشته می‌شوند تا معتبر به نظر برسند. با استفاده موفقیت‌آمیز از شبکه‌های عصبی در کاربردهای مختلف طبقه‌بندی، در این مقاله، سیستمی FakeGAN را پیشنهاد شده که برای اولین بار شبکه‌های متخاصم مولد (GANs) برای یک کار طبقه‌بندی متن، به‌ویژه، شناسایی بررسی‌های فریبنده تقویت و اتخاذ می‌کند. شبکه متخاصم مولد (GAN) یک چارچوب امیدوارکننده برای تولید نمونه‌های با کیفیت بالا با توزیع مشابه مجموعه داده هدف است FakeGANاز GANبرای یادگیری توزیع بررسی های واقعی و فریبنده و ایجاد یک طبقه بندی نیمه نظارت شده با استفاده از توزیع های مربوطه استفاده می کند.یک GAN از دو مدل تشکیل شده است: یک مدل تولیدی Gکه سعی می‌کند توزیع داده‌ها را به تصویر بکشد، و یک مدل متمایز D که بین نمونه‌هایی که از داده‌های آموزشی یا مولد G تمایز می‌یابد.این دو مدل به طور همزمان آموزش داده می شوند، جایی که G سعی می کند متمایز کننده D را فریب دهد، در حالی که D تخمین احتمال خود را به حداکثر می رساند که آیا یک نمونه از داده های آموزشی می آید یا توسط ژنراتور تولید می شود. به طور خلاصه، این چارچوب مربوط به یک بازی دو نفره Minimax است.بر خلاف مدل‌های استاندارد GAN که دارای یک مدل GeneratorوDiscriminator هستند، FakeGANاز دو مدل تشخیص‌دهنده و یک مدل مولد استفاده می‌کند. مولد به عنوان یک عامل خط مشی تصادفی در یادگیری تقویتی (RL) مدل‌سازی می‌شود و متمایزکننده‌ها از الگوریتم جستجوی مونت کارلو برای تخمین و ارسال مقدار اقدام میانی به عنوان پاداش RL به مولد استفاده می‌کنند. ارائه مدل مولد با دو مدل تفکیک کننده، با یادگیری از هر دو توزیع بررسی های صادقانه و فریبنده، از مسئله فروپاشی مد جلوگیری می کند. در واقع، آزمایش‌های ما نشان می‌دهد که استفاده از دو متمایزکننده، ثبات بالایی را برای FakeGAN فراهم می‌کند، که یک مسئله شناخته شده برای معماری‌های GAN است.در حالی که FakeGAN بر اساس یک طبقه‌بندی نیمه نظارت شده ساخته شده است، که به دلیل دقت کمتر شناخته شده است. ارزیابی انجام شده با استفاده از مجموعه داده‌ای متشکل از 800 بررسی از 20 هتل شیکاگو تریپ ادوایزر نشان می‌دهد که FakeGAN با دقت 89.1 درصد هم تراز با مدل‌های پیشرفته است. FakeGAN اولین گام خوبی را به سمت استفاده از GAN برای وظایف طبقه‌بندی متن، به ویژه آنهایی که به مجموعه داده‌های حقیقت زمینی بسیار بزرگ نیاز دارند، نشان می‌دهد.