



توسعه یک روش مبتنی بر CGAN برای تولید متن در سطح جنبه: استفاده از عوامل تشویق و تنبیه در دانش جنبه

محمد رضا شمس آ*، مریم لطفی شهرضا^ب، امیرمسعود سلطانی^ب

آ گروه مهندسی کامپیوتر، مرکز آموزش عالی شهرضا، دانشگاه اصفهان، ایران.
ب گروه هوش مصنوعی، دانشکده مهندسی کامپیوتر، دانشگاه اصفهان، ایران.



چکیده

سیستم‌های متن‌کاوی از تولید خودکار متن به‌ویژه زمانی که با مجموعه داده و منابع زبانی محدود سروکار دارند استفاده می‌کنند. اکثر رویکردهای موفق تولید متن، عمومی هستند و برای سطح جنبه ایجاد نشده‌اند، از این رو این روش‌ها، جملات نادرست و مشابه در جنبه‌های مختلف ایجاد می‌کنند. مطالعه حاضر با استخراج دانش جنبه از موضوعات مرتبط و ایجاد عبارت صحیح برای هر جنبه بر اساس شبکه متخاصم مولد مشروط (CGAN) راه حلی را برای این مشکل پیشنهاد می‌کند. روش پیشنهادی جملاتی را با استفاده از یک مجموعه داده کمکی تولید می‌کند که تمایزدهنده در تشخیص آن‌ها از جملات واقعی گمراه می‌شود. به منظور تولید این مجموعه داده کمکی، اطلاعات مبتنی بر جنبه از مجموعه داده‌های مرتبط با جنبه هدف استخراج می‌شود. برای بهبود بیشتر دقت، بسته به شباهت با مجموعه آموزشی، تولیدکننده تشویق یا تنبیه می‌شود. برای ارزیابی عملکرد روش تولید متن پیشنهادی از دو مجموعه داده به زبان انگلیسی و فارسی استفاده شده است. نتایج نشان می‌دهد که افزودن جنبه‌های مشابه به مجموعه داده کمکی کیفیت جملات تولید شده را بهبود می‌بخشد. علاوه بر این، تشویق منجر به جملات دقیق‌تر می‌شود، در حالی که تنبیه، تنوع جملات تولیدی را بالا می‌برد.
© 2023. تمامی حقوق محفوظ است.

اطلاعات مقاله

تاریخچه مقاله:
دریافت: 5 October 2022
اصلاح: 8 March 2023
پذیرش: 3 April 2023
انتشار آنلاین: 20 June 2023

کلمات کلیدی:

یادگیری عمیق، تولید متن، شبکه متخاصم مولد مشروط، جنبه، حافظه‌ی طولانی کوتاه مدت.

* نویسنده مسئول.

آدرس‌های رایانامه: m.r.shams@shr.ui.ac.ir (م. شمس)،
a.m.soltani@eng.ui.ac.ir (م. لطفی)، m.lotfi@shr.ui.ac.ir (ا. سلطانی)

تمامی حقوق محفوظ است. © 2023 ISSN: 2322-4460

