



Text Style Transfer in Persian Language using Deep Learning

Reza Khan Mohammadi and Seyed Abolghasem Mirroshandel

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

December 2, 2020

انتقال سبک متن در زبان فارسی با استفاده از یادگیری عمیق

رضا خان محمدی^۱، سید ابوالقاسم میرروشندل^۲

^۱ دانشجو کارشناسی گروه مهندسی کامپیوتر دانشکده فنی، دانشگاه گیلان
rezanecessaey@gmail.com

^۲ دانشیار و عضو هیئت علمی گروه مهندسی کامپیوتر دانشکده فنی، دانشگاه گیلان
mirroshandel@guilan.ac.ir

چکیده

در سال های اخیر، یادگیری عمیق طیف گسترده ای از امور پردازش زبان طبیعی را در بر گرفته است که از جمله به روز ترین آنها انتقال سبک متن می باشد. هدف این امر، تغییر نوشتار یک متن از سبک مبداء به سبک هدف، با حفظ محتوا و سلاست متن است. اکثر تحقیقات صورت گرفته این حوزه در زبان انگلیسی انجام شده و زبان فارسی به دلیل پیچیدگی زبانی، کمبود داده، و قدرت محاسباتی از تحولات موبطه به دور بوده است. در این پژوهش، نخستین روش انتقال سبک متن در دامنه جنسیت ارائه شده، و تاثیر کلمات و نقش های مختلف آنان بر تشخیص جنسیت متون مورد بررسی قرار گرفته است. به منظور سنجش میزان موفقیت روش های ارائه شده، ابتدا به تعلیم شبکه عصبی با هدف تشخیص جنسیت متن پرداخته و سپس با انتقال سبک و سعی بر شکست آن، نتایج بدست آمده را مقایسه می کنیم. قدرتمندترین روش پیشنهادی این پژوهش، موفق به شکست ۳۲.۹ درصدی مدل تشخیص جنسیت شده است.

کلمات کلیدی

انتقال سبک متن^۱، تشخیص جنسیت^۲، طبقه بندی کلمات^۳، جستجوی ابتکاری^۴، تعبیه سازی کلمات^۵، پردازش زبان طبیعی^۶، یادگیری عمیق^۷

۱- مقدمه

هدف این مقاله ارائه اولین روش انتقال سبک متن فارسی در حوزه جنسیت است. در این پژوهش، داده های متنی بکار گرفته شده [۱] به دو کلاس زن و مرد، که نمایانگر جنسیت نویسنده آنان است، طبقه بندی می شوند. سپس با اعمال روش پیشنهادی بر روی این متون، در پی تغییر سبک جنسیتی متن ورودی از جنسی به جنس دیگر هستیم. بدین منظور با کمک یادگیری عمیق به آموزش دو مدل تشخیص جنسیت و طبقه بندی کلمات می پردازیم. سپس با جانشینی کلمات در اجزای از پیش تعیین شده کلام^۱، استفاده از الگوریتم جستجوی ابتکاری و امتیازدهی بر اساس معادله معرفی شده، به تغییر سبک می پردازیم. حفظ معنا و سلاست دو معیار و چالش اساسی در سنجش کیفیت متون تولید شده توسط روش های انتقال سبک اند. لذا روش پیشنهادی با استفاده از

انتقال سبک^۱ یکی از امور نوظهور یادگیری عمیق است که در ابتدا بر روی داده های تصویری [4-2] اعمال شد. سپس با پیدایش معیار های مناسب دستی و خودکار ارزیابی، پیکره های متنی موازی، و مدل های زبانی از پیش آموزش دیده^۲، اعمال این امر بر داده های متنی [5, 6] به زبان انگلیسی رونق گرفت. پردازش زبان فارسی در سال های اخیر پیشرفت چشم گیری را به خود دیده، هرچند که تا به حال در زمینه انتقال سبک متن آزمایشی صورت نگرفته است، که از جمله دلایل این امر می توان به عدم وجود پیکره و ابزار های مناسب پردازشی در زبان فارسی اشاره کرد.

روش انتقال سبک بوده و شکست مدل پایه در شناسایی جنسیت نویسنده، به منزله موفقیت روش ارائه شده می‌باشد.

۴- مدل تشخیص جنسیت پایه

علی‌رغم کاربرد اصلی شبکه‌های عصبی پیچشی در پردازش داده‌های تصویری، این شبکه در استخراج ویژگی^{۱۸} داده‌های متنی نیز عملکرد قابل توجهی از خود نشان داده است. در طبقه‌بندی متون براساس جنسیت نویسنده، از یک مدل^۳ مجرای پیچشی^{۱۹} استفاده شده تا متن هدف با کمک اندازه کرنل، در رزولوشن‌های مختلف^۴، e و 8 پردازش شود. متغیر بودن اندازه کرنل در لایه‌های پیچشی باعث می‌شود تا با در نظر گرفتن تکه‌های مختلف در هر گام زمانی^{۲۰}، درک بهتری از بطن متن پیدا کنیم.

در ساختار این مدل علاوه بر لایه‌های پیچشی، از لایه‌های حافظه کوتاه‌مدت بلند^{۲۱} نیز استفاده شده است. مکانیسم درونی این لایه از سلول‌های حالت^{۲۲} و دروازه‌های فراموشی^{۲۳}، بروزسانی^{۲۴}، و خروجی^{۲۵} تشکیل شده و مدل را در راستای بخاطر سپردن ویژگی‌های استخراج شده توسط لایه‌های پیچشی یاری می‌کند.

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \quad (1)$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \quad (2)$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \quad (3)$$

$$\tilde{c}_t = \sigma_g(W_c x_t + U_c h_{t-1} + b_c) \quad (4)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \quad (5)$$

$$h_t = o_t \circ \sigma_h(c_t) \quad (6)$$

همانطور که نمایش داده شده، f_t ، i_t ، o_t هر کدام نمایانگر دروازه‌های فراموشی، بروزسانی، و خروجی لایه حافظه کوتاه‌مدت بلند هستند.

۵- روش پیشنهادی

طبق تحقیقات انجام شده در حوزه زبان‌شناسی اجتماعی، تفاوت‌های جنسیتی متون در انتخاب صفت و قید به حد اکثر می‌رسد. این پژوهش بر مبنای این تفاوت‌ها واقع شده و سعی بر آن دارد تا با شناسایی اجزای کلام (صفت، قید، فعل و ...)، مترادف‌یابی و متمایز کردن جنسیت آنان، و نهایتاً استخراج ترکیبی سلیس، سبک متون را تغییر دهد.

۵-۱- شناسایی اجزای کلام

در اولین مرحله از روش ارائه شده، اجزای مختلف کلام یک سند را با استفاده از کتابخانه پارسیوار [16] استخراج می‌کنیم. پرتکرارترین این اجزا عبارتند از: اسم، فعل، صفت، قید، ضمیر، اضافه، و حروف ربط. در طی آزمایشات انجام شده، تاثیر به کارگیری اجزای مختلف بررسی شده که در بخش ششم به آنان خواهیم پرداخت.

۵-۲- مترادف‌یابی و تفکیک جنسیت

داده‌های متنی از جمله داده‌های غیر ساختاریافته محسوب می‌شوند که برای پردازش، نیازمند انسجام اند. بسته کلمات ساختاریافتگی این داده‌ها را در غالب نمایش عددی کلمات ممکن ساخته، و با نگاشت کلمات متشابه به فضاهای برداری^{۲۶} متشابه، نمایشی Π -بعدی از هر کلمه در اختیار ما قرار می‌دهند. به منظور یافتن کلمات هم معنی از تشابه

بردار تعبیه کلمات^{۱۱} و جانیشینی کلمه‌ای با مترادفش، مفهوم جمله را حفظ کرده و با کمک معادله‌ای از ویژگی‌های دوتایی^{۱۲} و سه‌تایی^{۱۳} واژگان، از سلیس بودن جملات تولید شده اطمینان حاصل می‌کند.

بخش دوم مقاله مروری بر کارهای از پیش انجام شده در حوزه‌های تفاوت جنسیت^۴، طبقه‌بندی متون، و انتقال سبک خواهیم داشت. بخش سوم به تبیین مسئله پرداخته و بخش‌های چهارم و پنجم به ترتیب دربرگیرنده مدل پایه و روش معرفی شده است. در بخش ششم به آزمایشات انجام شده و نتایج اکتساب شده پرداخته شده، و در نهایت در بخش هفتم مقاله نتیجه‌گیری ذکر شده است.

۲- پیشینه و کارهای مرتبط

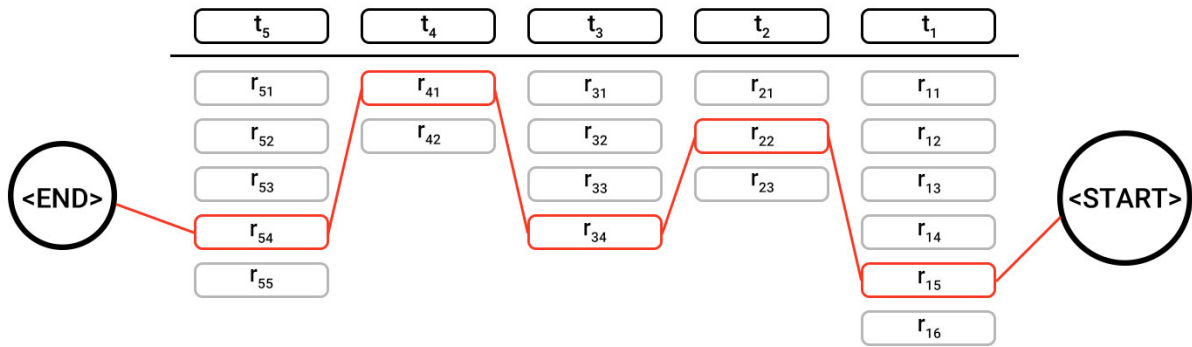
مبنای اصلی این مقاله، تفاوت‌های جنسیتی در متن می‌باشد که در زبان‌شناسی اجتماعی^{۱۵} به‌وفور به آن پرداخته شده. شناسایی نقش کلمات استفاده شده در برجسب‌های گفتار متفاوت در تعیین جنسیت متن [7-9]، از جمله مهمترین مباحث مورد تحقیق سال‌های اخیر زبان‌شناسی است.

در میان نخستین مدل‌های طبقه‌بندی متون فارسی، مرجع [10] با استفاده از مدل‌سازی موضوعی^{۱۶} توانست مفاهیم نهان میان کلمات را شناسایی کرده و مشکل ازدیاد ویژگی‌های استخراج شده حاصل از بکارگیری بسته کلمات را حل کند. از طرفی دیگر، در زمینه تشخیص جنسیت نویسنده در زبان فارسی، مرجع [۱] با بکارگیری مدل‌های آماری و یادگیری ماشین همچون بیز ساده^{۱۷}، درخت تصمیم تناوبی^{۱۸} و ماشین بردار پشتیبان^{۱۹} به طبقه‌بندی جنسیتی متون پرداخت.

از جمله روش‌های معرفی شده در زمینه انتقال سبک متون انگلیسی می‌توان به مراجع [11, 12, 13] اشاره نمود که به ترتیب با معرفی روش‌های مبتنی بر ترجمه ماشینی^{۲۰}، شبکه‌های خنمانه مولد^{۲۱}، یادگیری تقویتی^{۲۲}، و مدل زبانی مولد [5] به نحوی در این حوزه مشارکت داشته‌اند. از جمله مهمترین چالش‌های پیش روی این دسته از آزمایشات، کمبود داده‌های سبک هدف بوده که مرجع [14] با بکارگیری روش انطباق دامنه^{۲۳} بر این مشکل غلبه کرد. در حالی که اغلب پژوهش‌های انجام شده انتقال سبک متن قصد بر حفظ حداکثری مفاهیم با استفاده از مدل‌های پیچیده دارند، مرجع [15] با کمک بسته کلمات روشی ساده ولی کارآمد در این راستا ارائه داد که ما را در انجام این پژوهش یاری نمود.

۳- تبیین مسئله

پس از پیش پردازش متون غیرموازی و حذف نشانه‌ها و کلمات توقف، پیکره را به سه زیر مجموعه یادگیری^{۲۴}، اعتبارسنجی^{۲۵} و تست^{۲۶} تقسیم می‌کنیم و در ادامه با کمک زیرمجموعه یادگیری، به آموزش یک مدل شبکه عصبی^{۲۷} پایه به منظور طبقه‌بندی متون بر اساس جنسیت نویسنده آنان می‌پردازیم. سپس روش انتقال سبک را بر روی داده‌های تستی که مدل پایه قادر به پیش‌بینی درست آنان بوده اعمال می‌کنیم (متونی که به درستی مرد/زن پیش‌بینی شده بودند را به زن/مرد تبدیل می‌کنیم) و باری دیگر از مدل پایه به منظور پیش‌بینی متون منتقل شده (که در ابتدا به درستی پیش‌بینی شده بودند) می‌پردازیم. نتایج بدست آمده هرچه ضعیف‌تر باشند، نمایانگر تأثیرگذاری بیشتر



شکل (۱): نمایی کلی از روش پیشنهادی.

نکته: مثال بالا برای متنی با ۵ جزء اصلی t_{1-5} ذکر شده که هر کدام شامل مجموعه‌ی r_{ij} مربوط به خود اند. (هر مجموعه r_{ij} محدود به حداکثر ۱۰ کلمه هم‌معنی بوده. و از طرفی دیگر r_{ij} نمایانگر تعداد کلمات هم‌معنی برازنده سبک هدف است.)

الگوریتم‌های جستجوی ابتکاری، با انتخاب عناصر بهینه در هر گام، ما را در استخراج ترکیبی سلیس از میان ترکیب‌های موجود یاری می‌کند.

با ثبت تمامی دوتایی (b) ها و سه‌تایی (t) های کلمات مجاور موجود در پیکره، بهینگی انتخاب این الگوریتم را با تعریف تابع هیوریستیک زیر مشخص می‌کنیم.

$$h = \frac{6 \times t_i + 4 \times b_i}{10} \quad (7)$$

طبق فرمول (۷)، ابتکار الگوریتم جستجوی حریصانه در هر گام زمانی، به ترتیب ۴۰ و ۶۰ درصد وابسته به تعداد دفعات همجواری کلمه‌ای با یک (b_i) و دو (t_i) کلمه آخر مسیر پیمایش شده است. از طرفی دیگر، با در نظر گرفتن دو توکن <START> و <END> به عنوان نقاط ثابت شروع و پایان، شرایطی منسجم‌تر برای جستجو فراهم می‌کنیم. شکل (۱) نمایانگر شیوه جستجوی الگوریتم پیشنهادی در میان کاندیداهای جایگزینی اجزای انتخاب‌شده t_{1-5} است.

۶- نتایج

در این پژوهش از پیکره متنی دارای برچسب جنسیت [۱] استفاده شده که در هر یک از دو جنس، شامل ۲۴۵ سند مختلف به زبان فارسی رسمی است. اغلب این اسناد، کتاب و رمان‌های فارسی هستند که توسط نویسندگان مرد یا زن نوشته شده‌اند. از آنجایی که طول هر یک از اسناد متفاوت بوده، هر سند را به اجزای کوچکتری تقسیم می‌کنیم. اطلاعاتی تکمیلی را در جدول (۱) مشاهده می‌کنیم.

جدول (۱): مقایسه پیکره متنی

جنسیت	یادگیری	اعتبارسنجی	تست	مجموع
مرد	۵,۱۲۰	۱,۳۳۹	۱,۵۹۱	۸,۰۴۰
زن	۸,۴۵۸	۲,۰۶۶	۲,۶۵۳	۱۳,۱۷۷

کسینوسی، یکی از معروف‌ترین مقیاس‌های اندازه‌گیری همانندی دو بردار، استفاده می‌کنیم.

$$similarity = \cos(\theta) = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (6)$$

پس از تشخیص اجزای کلام مورد نظر، در مرحله دوم با اختصاص A به بردار هر جزء و B به بردار دیگر کلمات حاضر در دایره واژگان^{۳۷}، مجموعه‌ای از کلمات هم‌معنی برای هر جزء به عنوان کاندیداهای جایگزینی بدست خواهد آمد. حال به منظور انتقال سبک، نیازمند آنیم تا از میان کاندیداهای کلماتی را که برازنده سبک مبدأ/هدف هستند را از دامنه انتخابات حذف/انتخاب کنیم.

در مرحله سوم به آموزش مدلی دیگر از شبکه‌های عصبی، متشکل از لایه‌های پیمایشی و حافظه کوتاه‌مدت بلند می‌پردازیم. اینبار قصد داریم تا کلمات را از لحاظ جنسیتی تفکیک کنیم. لذا با استخراج صفات، افعال، اسامی و قیده‌های موجود در پیکره متنی موجود از هر جنس، داده‌های آموزشی لازم را فراهم می‌کنیم. کمبود اطلاعات کافی به منظور استخراج ویژگی، همواره طبقه‌بندی کلمات را برخلاف اسناد و جملات به امری دشوار تبدیل نموده است. لذا با پایین آوردن تراز پردازش متون از کلمه به کاراکتر، ویژگی‌های قابل استخراج بیشتری در دست خواهیم داشت.

در نهایت با طبقه‌بندی کلمات پیشنهادی توسط مدل معرفی شده، به ازای هر جزء مورد نظر، مجموعه‌ای از کاندیداهای جایگزینی داریم که (۱) هم‌معنی و (۲) به سبک مخالف اند.

۳-۵- استخراج ترکیبی سلیس

همانطور که در بخش‌های قبلی به آن اشاره شد، سلاست متون منتقل شده از معیارهای اساسی ارزشیابی کیفی روش‌های پیشنهادی این حوزه است. بدین منظور نیاز است تا روشی مناسب در استخراج متن نهایی اتخاذ شود.

از الگوریتم جستجوی حریصانه^{۳۸} می‌توان به عنوان یکی از مهمترین رمزگشاهای^{۳۹} موجود در حوزه ترجمه ماشینی یاد کرد. این عضو از خانواده‌ی

جدول (۳): تاثیر انواع روش های انتقال سبک بر تضعیف مدل تشخیص جنسیت

توازن	H	F	دقت مدل پایه			روش
			T_w	T_r	T_a	
۲.۸۵	۶۲	۱۸۳	۱۴.۶۹	۹۵.۲۲	۸۷.۳۴	مبتنی بر کلمه + (صفت، قید)
۴.۸۵	۱۱۲	۳۱۶	۲۶.۹۳	۹۱.۷۴	۸۵.۳۹	مبتنی بر کلمه + (صفت، قید، فعل)
۲۸.۶۵	۱۴۷	۱,۳۶۳	۳۵.۲۸	۶۴.۳۹	۶۱.۵۴	مبتنی بر کلمه + (صفت، قید، فعل، اسم)
۳۳.۱۵	۱۵۶	۱,۵۶۳	۳۷.۴۰	۵۹.۱۵	۵۷.۰۴	مبتنی بر کاراکتر + (صفت، قید، فعل، اسم)

(۲) تغییر تراز: بصورت کلی تراز پردازشی مدل های تفکیک جنسیت هر روش ارائه شده مبتنی بر یکی از دو تراز کلمه یا کاراکتر است (شکاف ایجاد شده بین روش های ۳ و ۴).

درصدهای کسب شده در مجموعه های T_a ، T_r و T_w به معنای تاثیرگذاری کلی، گمراهی مدل، و کمک ناخواسته اند. شیب نزولی دقت مدل پایه در مجموعه تست به منزله افزایش تاثیرگذاری مدل های ۱ تا ۴ است. از طرفی دیگر، با تضعیف مدل پایه در زیرمجموعه T_r ، قادر به گمراه کردن مدل در تشخیص جنسیت F داده ای هستیم که قبل از انتقال بدرستی تشخیص داده شده بودند. سرآخر، درصدهای ارائه شده در ستون T_w نمایانگر میزان خطای هر روش است، بدین صورت که به طور ناخواسته باعث شده تا مدل در H تعداد داده که توانایی طبقه بندی درست آن ها را نداشته، صحیح عمل کند. به منظور مقایسه درست موفقیت روش های ارائه شده، مقیاس زیر را معرفی کرده تا توازنی میان دو مقدار F و H برقرار کنیم.

$$\text{توازن} = \frac{F-H}{n} \times 100 \quad (۸)$$

همانطور که قابل مشاهده است، با کوچکتر شدن تراز پردازشی و افزایش دامنه، تاثیرگذاری روش ها، در قلب مقادیر توازن بالاتر، افزایش یافته. روش نهایی این پژوهش (روش شماره ۴) موفق به کاهش دقت مدل پایه از ۹۰٪ به ۵۷.۰۴٪ شده است.

۷- نتیجه گیری

انتقال سبک متن یکی از بروزترین امور پردازش زبان طبیعی بوده که نقش یادگیری عمیق در آن بسیار است. در این مقاله، با تکیه بر تفاوت های زبان شناختی اجتماعی. اولین نمونه فارسی این روش ها ارائه شد. حفظ معنا و سلاست متون از جمله چالش های پیش رو این دسته از تحقیقات بوده که در روش های معرفی شده برترتیب با استفاده از بسته کلمات و تابع هیوربستیک مناسب برطرف شدند. سرآخر با کاهش دقت مدل تشخیص جنسیت به مقدار ۳۲.۹٪، گامی در جهت پیشرفت های آتی این حوزه از پردازش زبان فارسی برداشته شد.

مدل های آزمایش شده در این پیکره متنی شامل بیز ساده، رگرسیون لجستیک، برت چندزبانه، و مدل مطرح شده در بخش ۴ (CNN + LSTM NN) است. مدل نهایی ارائه شده در این پژوهش موفق شد تا ۱۸٪ پیشرفته ترین دقت از پیش ارائه شده را بهبود دهد.

جدول (۲): مقایسه مدل ها

دقت	مدل
۶۲٪	رگرسیون لجستیک
۶۹٪	بیز ساده
۷۰٪	برت چندزبانه
۷۲٪	ماشین بردار پشتیبان
۹۰٪	CNN + LSTM NN

حال به منظور بررسی میزان موفقیت روش انتقال سبک، نمادهای S_t و S_r را به ترتیب به سبک مبدأ، سبک هدف، و مجموعه تست منتقل نشده مدل پایه اختصاص می دهیم. سپس با تقسیم D_a به دو زیرمجموعه D_r و D_w ، به عنوان مجموعه داده های تستی که مدل پایه آن ها را به ترتیب درست و غلط پیش بینی کرده بود، مجموعه های تست را منتقل می کنیم و آن ها را T_a ، T_r و T_w نامگذاری می کنیم.

همانطور که در جدول (۳) قابل مشاهده است، نام هر روش تعیین کننده تراز پردازشی مدل تفکیک جنسیت و مجموعه اجزای کلام آن است که هر دو رابطه مستقیمی با تضعیف مدل پایه دارند. بدیهیست که با طبقه بندی هر کدام از مجموعه های D_a ، D_r و D_w (شامل ۴۲۴۴، ۳۸۲۸، و ۴۱۶ داده) به دقت ۹۰، ۱۰۰، و ۰ درصد دست پیدا می کنیم. پس از انتقال متون، شاهد دو الگوی تغییر متفاوت در آزمایش های انجام شده هستیم:

(۱) تغییر دامنه: با افزایش تعداد اجزای کلام مورد نظر، تغییرات هر روش بر دامنه بیشتری از اجزای اسناد ورودی اعمال شده که همواره باعث تضعیف مدل پایه شده است (روش های ۱-۳).

- [10] Ahmadi, P., Tabandeh, M., and Gholampour, I., Persian text classification based on topic models, ۲۰۱۶.
- [11] Subramanian, Sandeep, Lample, Guillaume, Michael Smith, Eric, Denoyer, Ludovic, Ranzato, Marc'Aurelio, Boureau, Y-Lan, Multiple-Attribute Text Style Transfer, ۲۰۱۹.
- [12] Yang, Zichao, Hu, Zhiting, Dyer, Chris, Xing, Eric P., Berg-Kirkpatrick, Taylor, Unsupervised Text Style Transfer using Language Models as Discriminators, ۲۰۱۹.
- [13] Gong, Hongyu, Bhat, Suma, Wu, Lingfei, Xiong, Jinjun, Hwu, Wen-mei, Reinforcement learning based text style transfer without parallel training corpus, ۲۰۱۹.
- [14] Li, Dianqi, Zhang, Yizhe, Gan, Zhe, Cheng, Yu, Brockett, Chris, Sun, Ming-Ting, Dolan, Bill, Domain adaptive text style transfer, ۲۰۱۹.
- [15] John, Vineet, Mou, Lili, Bahuleyan, Hareesh, Vechtomova, Olga, Disentangled Representation Learning for Non-Parallel Text Style Transfer, ۲۰۱۸.
- [16] Mohtaj, Salar, Roshanfekr, Behna, Zafarian, Atefeh, Asghari, Habibollah, Parsivar: A Language Processing Toolkit for Persian, Proceedings of the Eleventh International Conference on Language Resources and Evaluation, ۲۰۱۸.
- [1] مرادی، مهدی، بحرانی، محمد، تشخیص خودکار جنسیت نویسنده در متون فارسی، ۱۳۹۴.
- [2] Gatys, Leon, Ecker, Alexander, Bethge, Matthias, Preserving Color in Neural Artistic Style Transfer, ۲۰۱۵.
- [3] Gatys, Leon, Bethge, Matthias, Hertzmann, Aaron, Shechtman, Eli, Preserving Color in Neural Artistic Style Transfer, ۲۰۱۶.
- [4] Luan, Fujun, Paris, Sylvain, Shechtman, Eli, Bala, Kavita, Preserving Color in Neural Artistic Style Transfer, ۲۰۱۷.
- [5] Sudhakar, Akhilesh, Upadhyay, Bhargav, Maheswaran, Arjun, Transforming Delete, Retrieve, Generate Approach for Controlled Text Style Transfer, ۲۰۱۹.
- [6] Lee, Joseph, Xie, Ziang, Wang, Cindy, Drach, Max, Jurafsky, Dan, Ng, Andrew Y., Neural Text Style Transfer via Denoising and Reranking, ۲۰۱۹.
- [7] Hoyle, Alexander, Wolf-Sonkin, Wallach, Hanna, Augenstein, Isabelle, Cotterell, Ryan, Unsupervised Discovery of Gendered Language through Latent-Variable Modeling, ۲۰۱۹.
- [8] Kreyer, Rolf, Review of Baker (2014): Using Corpora to Analyze Gender, ۲۰۱۴.
- [9] Pearce, Michael, Investigating the collocational behaviour of MAN and WOMAN in the BNC using Sketch Engine, ۲۰۰۸.

Generative Adversarial Network	۲۱
p reinforcement learning	۲۲
Domain Adaptation	۲۳
Train set	۲۴
Validation set	۲۵
Test set	۲۶
Neural Network	۲۷
Feature Extraction	۲۸
Convolutional	۲۹
Time step	۳۰
Long short-term memory (LSTM)	۳۱
State Cell	۳۲
Forget Gate	۳۳
Update Gate	۳۴
Output Gate	۳۵
Vector Spaces	۳۶
Vocabulary	۳۷
Greedy Search Algorithm	۳۸
Decoder	۳۹

Text Style Transfer	۱
Gender Identification	۲
Word Classification	۳
Heuristic Search	۴
Bag of Words	۵
Natural Language Processing	۶
Deep Learning	۷
Style Transfer	۸
Pre-trained Language Models	۹
Part of Speech Tag	۱۰
Word Embedding	۱۱
Bigram	۱۲
Trigram	۱۳
Gender Difference	۱۴
Sociolinguistics	۱۵
Topic Modeling	۱۶
Naïve Bayes	۱۷
Alternating decision tree	۱۸
Support vector machine	۱۹
Machine Translation	۲۰