علوم رایانش و فناوری اطلاعات
نشریه علمی انجمن کامپیوتر ایران
مجلد ۱۷، شماره ۲، ۱۳۹۸
صفحات ۴۴- ۳۳
مقاله عادی (مروری)

یادگیری دوگان ژرف و کاربردهای آن

علی‌اکبر خوش‌ویشکائی[۱]، حمید بیگی[۲]*

*نویسنده مسئول، دریافت: ۹۸/۰۸/۲۳، بازنگری: ۹۸/۱۱/۰۵، پذیرش: ۹۹/۰۲/۲۹

[۱] دانش‌آموخته کارشناسی ارشد، مهندسی کامپیوتر، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف، تهران، ایران
[۲] دانشیار، مهندسی کامپیوتر، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف، تهران، ایران

چکیده

یادگیری ژرف، که در بسیاری از مسائل هوش مصنوعی به نتایج بسیار خوبی رسیده است، از این واقعیت رنج می‌برد که عملکرد آن به‌شدت به حجم داده‌های برچسب‌دار بستگی دارد. در بسیاری از کاربردهای دنیای واقعی، تعداد نمونه‌های دارای برچسب معمولاً محدود بوده و گردآوری آن نیز پرهزینه است. درحالی‌که اغلب، نمونه‌های بدون برچسب به مقدار کافی موجود است. بنابراین، ارائه‌ی روش‌هایی برای بهره‌برداری مؤثر از نمونه‌های بدون برچسب توجه بسیاری را به خود جلب کرده است. گذشته از این، بسیاری از مسائل هوش مصنوعی در قالب دوگان ظاهر می‌شوند؛ برای نمونه، ترجمه انگلیسی به فارسی در مقابل ترجمه فارسی به انگلیسی و طبقه‌بندی تصویر در مقابل تولید تصویر. در سال‌های اخیر، روش‌های متعددی برای استفاده از همبستگی بین وظایف دوگان ارائه شده است. در این مقاله، به بررسی روش‌های یادگیری دوگان می‌پردازیم، که هدف از آن بهره‌برداری مؤثر از دوگانگی میان دو وظیفه‌ی دوگان در آموزش و یا استنتاج است. یادگیری دوگان را می‌توان به سه سطح مختلف، یعنی دوگانگی در سطح داده، در سطح مدل و در سطح استنتاج تقسیم نمود. در این مقاله، به روش‌های مختلف برای بهره‌گیری از این ایده‌ها و موفقیت‌های آن‌ها در کاربردهای مختلف، خواهیم پرداخت. همچنین نشان خواهیم داد که چگونه یادگیری دوگان به‌طور مؤثر نیاز به داده‌های دارای برچسب را کاهش می‌دهد.

کلمات کلیدی: یادگیری ژرف، یادگیری دوگان، وظایف دوگان

۱- مقدمه

یادگیری ژرف زیرمجموعه‌ای از یادگیری ماشین در هوش مصنوعی است که به مدل‌های محاسباتی تشکیل شده از چندین لایه‌ی پردازشی اجازه می‌دهد تا بازنمایی از داده‌ها با سطوح انتزاع مختلفی را یاد بگیرند [۱]. بازنمایی هر لایه از طریق مقادیر بازنمایی در لایه‌ی قبلی و وزن‌های لایه‌ها (پارامترهای مدل) به دست می‌آید. یادگیری ژرف با استفاده از الگوریتم پس‌انتشار[۱] ساختارهای پیچیده‌ای را در مجموعه مجموعه داده‌های بزرگ کشف می‌کند تا نشان دهد چگونه یک ماشین باید پارامترهای داخلی خود به‌روزرسانی نماید [۱]. از یادگیری ژرف ، به‌جرئت می‌توان به‌عنوان یکی از هیجان‌انگیزترین فناوری‌های دهه‌ی اخیر یاد کرد. در سال‌های اخیر شاهد نتایج خیره‌کننده‌ی آن در بسیاری از مسائل مانند ترجمه‌ی ماشینی [۲]، دسته‌بندی تصاویر [۳]، تولید تصاویر [۱]، تشخیص گفتار [۵، ۶] و تولید گفتار [۷] بوده‌ایم. این نتایج توجه بسیاری را به خود جلب کرده است، به‌گونه‌ای که امروزه شاهد استفاده از یادگیری ژرف در اغلب زمینه‌ها هستیم.

اگرچه یادگیری ژرف توانسته است بهبود قابل‌ملاحظه‌ای در حل این مسائل به دست آورد، همچنان به داده‌های برچسب‌دار زیادی نیاز دارد. از طرفی گردآوری چنین مجموعه داده‌های بزرگی، اغلب کاری دشوار و بسیار پرهزینه است. در مقابل، مجموعه داده‌های بدون برچسب زیادی وجود دارد که می‌توان از آن‌ها جهت بهبود نتایج و آموزش بیشتر بهره برد. ازاین‌رو، پژوهش‌های متعددی درزمینه‌ی یادگیری نیمه نظارتی و بدون نظارت و کاهش وابستگی آموزش به مجموعه داده‌های برچسب‌دار بزرگ صورت گرفته است.

از سوی دیگر، بسیاری از این مسائل به‌صورت دوگان[۲] ظاهر می‌شوند [۸، ۹]؛ یعنی دو مدل به‌گونه‌ای باهم در ارتباط می‌باشند که ورودی‌ها و خروجی‌های یکی به ترتیب خروجی‌ها و ورودی‌های دیگری است. مترجم ماشینی از زبان فارسی به انگلیسی و از زبان انگلیسی به فارسی، دسته‌بندی تصاویر و تولید تصاویر و درنهایت

تشخیص گفتار و تولید گفتار، مثال‌هایی چند از جفت مدل‌هایی به‌صورت دوگان است. درحالی‌که می‌توان از همبستگی موجود میان این مسائل استفاده نمود، در بیش‌تر روش‌های یادگیری این موضوع نادیده گرفته شده و هر کدام از این مدل‌ها به‌صورت جداگانه آموزش داده می‌شوند [۸].

یادگیری دوگان[۳]، تلاشی با هدف بهبود نتایج و کاهش نیاز به مجموعه داده‌های بزرگ برچسب‌دار از طریق آموزش همزمان دو مدل دوگان است. همواره در یادگیری دوگان به دو مسئله به‌طور همزمان پرداخته می‌شود که به یکی از آن‌ها وظیفه‌ی اصلی[۴] و به دیگری وظیفه‌ی دوگان[۵] می‌گویند. به‌این‌ترتیب خروجی مدل وظیفه‌ی اصلی ورودی مدل وظیفه‌ی دوگان خواهد بود و برعکس. در حالت تعمیم‌یافته‌ی آن که یادگیری حلقه‌بسته[۶] نام دارد، چندین وظیفه به دنبال هم در یک مسیر بسته قرار می‌گیرند به‌طوری‌که خروجی هر مدل ورودی مدل بعدی باشد.

ایده‌ی یادگیری دوگان نخستین بار توسط تیم تحقیقاتی مایکروسافت در سال ۲۰۱۶ در آموزش مترجم ماشینی عصبی به کار گرفته شد و هدف از آن، آموزش دو مترجم به‌صورت همزمان بود به‌گونه‌ای که دو مدل به‌صورت همکار باهم تعامل داشته و سبب بهبود یکدیگر شوند [۸]. درواقع، دو مترجم ماشینی، یکی از زبان انگلیسی به فرانسوی و دیگری از زبان فرانسوی به انگلیسی، در تعامل با یکدیگر با استفاده از روش‌های گرادیان سیاست[۷] در یادگیری تقویتی آموزش داده شدند [۸]. به این صورت که جمله‌ای از زبان انگلیسی، ابتدا با استفاده از مدل اول به زبان فرانسوی ترجمه شده، سپس جمله‌ی حاصل با استفاده از مدل دوم به زبان انگلیسی بازگردد و انتظار می‌رود این دو جمله‌ی انگلیسی مشابه یکدیگر باشند.

به‌صورت شهودی، در حالت استاندارد یادگیری از روی مجموعه داده توزیع $P(y|x)$ را که در آن x داده‌ی ورودی و y برچسب متناظر است، مدل‌سازی می‌شود. اما یادگیری دوگان در کنار این توزیع، توزیع $P(x|y)$ را نیز مدل‌سازی می‌نماید و با تعامل دو مدل، با استفاده از بازخوردهای مؤثر یا یک عبارت منظم‌ساز[۸]، تلاش می‌کند تخمین بهتری برای هر دو توزیع به دست آورد. به‌این‌ترتیب، نه‌تنها نتایج بهتری با استفاده از این روش به دست می‌آید، نیاز به داده‌های برچسب‌دار جهت آموزش دو مدل نیز به‌شدت کاهش می‌یابد. در سه سال اخیر، پژوهش‌های متعددی در این زمینه صورت پذیرفته و این روش در مسائل گوناگونی از قبیل ترجمه‌ی تصاویر [۱۰، ۱۱]، تحلیل و تولید تمایل [۹]، تولید و پاسخ‌گویی به پرسش [۱۲] و همچنین تشخیص و تولید گفتار [۱۳] اعمال شد. توجه به رابطه‌ی دوگان میان وظایف به همین‌جا ختم نمی‌شود و از این همبستگی به گونه‌های دیگری، مانند دوگانگی در سطح معماری مدل[۹] [۱۴] و در سطح استنتاج[۱۰] [۱۵]، نیز بهره برده شده است. بر این اساس، می‌توان یادگیری دوگان را به سه سطح یادگیری دوگان در سطح داده، در سطح معماری مدل و در سطح استنتاج دسته‌بندی نمود. یادگیری دوگان در سطح داده خود می‌تواند شامل بخش‌های مختلفی ازجمله یادگیری بدون نظارت، یادگیری بانظارت، یادگیری نیمه‌نظارتی، یادگیری انتقالی و یادگیری تخاصمی[۱۱] گردد.

در این مقاله مروری بر پژوهش‌های صورت گرفته در زمینه‌ی یادگیری دوگان ژرف و کاربردهای آن خواهیم داشت. در ادامه ضمن بررسی روش یادگیری دوگان ژرف، یک دسته‌بندی برای این روش‌ها ارائه و سپس هر کدام از این دسته‌ها بررسی می‌گردند. در پایان کاربردهای مختلفی را که از یادگیری دوگان ژرف برای حل آن‌ها بهره برده شده است مرور خواهیم نمود. ادامه‌ی این مقاله به‌صورت زیر سازمان‌دهی شده است. نخست، در بخش ۲ مقدمه‌ای از یادگیری ژرف بیان می‌نماییم. سپس در بخش ۳ روش‌های یادگیری دوگان در سطح داده، سطح مدل و همچنین سطح استنتاج را بررسی خواهیم نمود. در بخش بعدی به کاربردهای یادگیری دوگان در مسائل مختلف، پژوهش‌های صورت گرفته در این زمینه و نتایج به‌دست‌آمده خواهیم پرداخت و درنهایت، در بخش ۵ با جمع‌بندی و نتیجه‌گیری مقاله را به پایان می‌رسانیم.

۲- یادگیری ژرف

یادگیری ژرف شاخه‌ای از یادگیری ماشین و مبتنی بر شبکه‌های عصبی مصنوعی است. شبکه‌های عصبی مصنوعی، همان‌طور که از نام آن برمی‌آید قصد دارد تا از شبکه‌های عصبی مغز انسان تقلید نماید. بنیادی‌ترین واحد یک شبکه عصبی ژرف، نورون مصنوعی[۱۲] نامیده می‌شود که ورودی را گرفته، آن را پردازش می‌کند، سپس آن را از طریق یک تابع فعالیت غیرخطی مانند سیگموئید[۱۳] عبور داده و مقدار حاصل را در خروجی باز می‌گرداند. این شبکه‌های مصنوعی از چندین لایه تشکیل شده است که تعداد این لایه‌ها عمق شبکه را نشان می‌دهد. هر لایه شامل چندین نورون است و نورون‌های هر لایه می‌تواند به همه و یا تعدادی از نورون‌های لایه‌ی بعدی متصل باشد. یک شبکه‌ی عصبی با تنها یک لایه‌ی نهان برای بازنمایی هر تابعی کافی است، اما این لایه می‌تواند بسیار بزرگ بوده و غیر قابل آموزش باشد [۱۶]. ازاین‌جهت به شبکه‌های عصبی مصنوعی تقریب‌گر فراگیر[۱۴] گفته می‌شود.

یکی از تفاوت‌های عمده‌ی یادگیری ژرف با روش‌های سنتی یادگیری ماشین، مانند دسته‌بند ماشین بردار پشتیبان، در مهندسی ویژگی‌ها است. برای مثال مسئله‌ی دسته‌بندی را در نظر بگیرید. دقت دسته‌بند به‌شدت تحت تأثیر ویژگی‌های استخراج شده از داده‌ها می‌باشد و درصورتی‌که ویژگی‌های در نظر گرفته شده مناسب نباشند، دسته‌بند عملکرد خوبی نخواهد داشت. در یادگیری ژرف، برخلاف روش‌های سنتی، ویژگی‌ها به‌صورت خودکار استخراج می‌گردد. در لایه‌های ابتدایی ویژگی‌های ساده‌ای از روی داده‌ی ورودی به‌دست‌آمده و در لایه‌های بعدی از روی ویژگی‌های به‌دست‌آمده در لایه‌ی قبلی ویژگی‌های پیچیده‌تری استخراج می‌گردد. در پایان، یک یا چند لایه‌ی نهایی همانند یک دسته‌بند بر روی آن ویژگی‌های نهایی به‌دست‌آمده عمل می‌کند.

وجود تعداد لایه‌های متعدد سبب می‌شود تا تعداد پارامترهای مدل در مقایسه با روش‌های سنتی به مقدار قابل‌توجهی بیش‌تر باشد. از طرفی، با افزایش تعداد پارامترها نیاز به مجموعه داده یا پیچیدگی نمونه‌ای[۱۵] افزایش می‌یابد که این موضوع یک خبر ناخوشایند برای یادگیری ژرف است. در سال‌های اخیر، با پیشرفت‌های متعدد در یادگیری ژرف، تلاش‌هایی جهت کاهش نیاز به مجموعه داده‌های برچسب‌دار صورت پذیرفته است.

آموزش شبکه‌های عصبی ژرف با استفاده از روش پس‌انتشار صورت می‌پذیرد. در این روش، ابتدا گراف محاسباتی شبکه ایجاد شده و داده‌ها به لایه‌ی ورودی داده می‌شود. سپس ورودی نورون‌ها در لایه‌های بعدی با جمع وزن‌دار نورون‌های لایه‌ی قبلی به دست می‌آید که این وزن‌ها، درواقع، پارامترهای شبکه می‌باشند. پس از به دست آمدن خروجی از لایه‌ی آخر، مقدار تابع زیان محاسبه می‌گردد. حال، گرادیان تابع زیان نسبت به پارامترهای لایه‌ی خروجی محاسبه شده و لایه به لایه در جهت عقب‌گرد انتشار می‌یابد. سپس با توجه به الگوریتم بهینه‌سازی مورد نظر پارامترهای شبکه به‌روزرسانی می‌شود. این کار ادامه می‌یابد تا پارامترهای مناسب شبکه به دست آید.

البته آموزش شبکه‌های ژرف همواره به این سادگی نبوده و با افزایش عمق و تعداد پارامترها چالش‌های جدیدی ازجمله انفجار یا ناپدید شدن گرادیان به وجود می‌آید. این شبکه‌ها نه‌تنها به مجموعه داده‌های بزرگ نیازمندند، بلکه فرآیند آموزشی با بار محاسباتی طاقت‌فرسا دارند و عمده‌ی موفقیت خود را مدیون مجموعه داده‌های بزرگ گردآوری شده در سال‌های اخیر و قدرت محاسباتی به‌دست‌آمده با استفاده از واحدهای پردازش گرافیکی[۱۶] هستند.

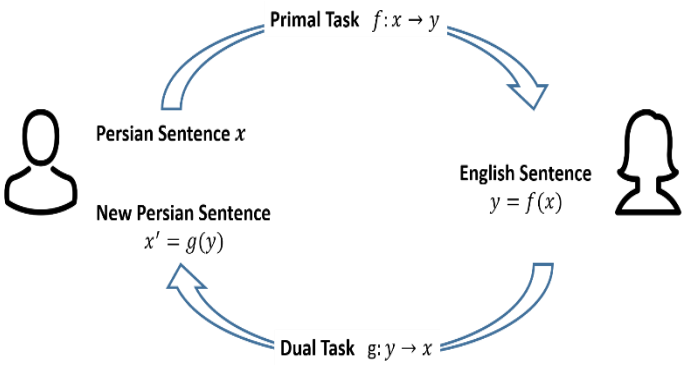
۳- یادگیری دوگان ژرف

به‌طورکلی، در مسائل یادگیری بانظارت با مجموعه داده‌ای به شکل $S = \{(x^{(i)}, y^{(i)})\}_{i=1}^{n}$ روبرو هستیم و سعی داریم مدلی را یاد بگیریم که به ازای

هر ورودی x خروجی y را به درستی پیش‌بینی نماید. به‌این‌ترتیب، قصد داریم توزیع $P(y|x;\theta)$ را که در آن θ مجموعه پارامترهای مدل است، فرا گیریم. در یادگیری ژرف، این توزیع توسط یک شبکه‌ی عصبی ژرف مدل‌سازی می‌شود که در آن وزن‌های لایه‌ها درواقع همان مجموعه پارامتر θ هستند که باید یاد گرفته شود. به‌این‌ترتیب با داشتن مجموعه داده‌ی مورد نظر و با استفاده از روش پس‌انتشار، طی تکرارهای[۱۷] متعدد پارامترهای مدل به‌روزرسانی شده تا جایی که مقادیر مناسبی برای این پارامترها به دست آید. به‌طور کلی، پارامترهای مدل در هر به‌روزرسانی به‌گونه‌ای تغییر می‌یابند که تابع زیان تعریف شده برای آن مسئله‌ی خاص کاهش یابد و درنهایت مناسب بودن پارامترها با استفاده از مجموعه داده‌ی آزمون ارزیابی می‌گردد.

در یادگیری دوگان، دو وظیفه به‌طور همزمان آموزش داده می‌شود که به یکی وظیفه‌ی اصلی و به دیگری وظیفه‌ی دوگان گفته می‌شود. مجموعه داده‌ی S را برای مسئله‌ی تشخیص گفتار در نظر بگیرید. به‌این‌ترتیب، هر $x^{(i)}$ یک سیگنال صوتی و $y^{(i)}$ عبارت متنی متناظر با آن است. اگر قصد حل این مسئله را با استفاده از روش یادگیری دوگان داشته باشیم، به مسئله‌ی تشخیص گفتار، وظیفه‌ی اصلی می‌گوییم که در آن قصد داریم تابع $f_{\theta_1}: x \to y$ با پارامترهای θ_1 را به دست آوریم. به وضوح، این وظیفه تابع توزیع $P(y|x;\theta_1)$ را در خود جای داده است. در عین حال، می‌توان از روی این مجموعه داده مدلی را برای تبدیل $y^{(i)}$ (عبارت متنی) به $x^{(i)}$ (سیگنال صوتی) نیز آموزش داد. در این حالت، به‌طور مشابه تابع $g_{\theta_2}: y \to x$ با مجموعه پارامتر θ_2 را به دست می‌آوریم که توزیع $P(x|y;\theta_2)$ را در خود جای داده است. این مسئله، تولید گفتار است که در این مثال به آن وظیفه‌ی دوگان می‌گوییم. بنابراین خروجی وظیفه‌ی اصلی ورودی وظیفه‌ی دوگان است و بر عکس. بسیاری از مسائل یادگیری ماشین می‌توانند در قالب دوگان[۱۸] قرار گیرند.

در یادگیری دوگان با آموزش همزمان دو مدل وظیفه‌ی اصلی و دوگان سعی بر بهبود نتایج و همچنین کاهش نیاز به مجموعه داده‌ی برچسب‌دار است. به‌طور کلی، این هدف با کاهش مجموع تعداد پارامترهای دو مدل یا افزایش ضمنی اندازه‌ی مجموعه داده صورت می‌پذیرد و به‌این‌ترتیب، دو مدل به قدرت تعمیم[۱۹] بیشتری دست می‌یابند. در این روش با دو دامنه‌ی داده‌ی X و Y سر و کار داریم که در وظیفه‌ی اصلی ورودی از دامنه‌ی X به خروجی در دامنه‌ی Y تبدیل شده و در وظیفه‌ی دوگان بر عکس. با توجه به این که این دو وظیفه بر عکس یکدیگرند، ورودی هر مدل را می‌توان از روی خروجی آن و با استفاده از مدل دیگر بازسازی نمود و ایده‌ی اصلی در یادگیری دوگان کاهش خطای بازسازی[۲۰] است. بهره‌گیری از دوگانگی موجود میان مسائل، می‌تواند در سطوح مختلفی ازجمله داده، معماری مدل و همچنین در استنتاج صورت گیرد. در این بخش مروری بر پژوهش‌های صورت گرفته در هر کدام از این سطوح خواهیم داشت.



شکل ۱- ترجمه‌ی ماشینی با استفاده از یادگیری دوگان

۱-۳- یادگیری دوگان در سطح داده

در یادگیری دوگان در سطح داده، همبستگی موجود میان وظایف اصلی و دوگان از طریق مجموعه داده استخراج می‌گردد. به‌طور کلی، این آموزش می‌تواند از طریق افزودن یک عبارت منظم‌ساز یا با استفاده از روش گرادیان سیاست صورت پذیرد. از عبارت منظم‌ساز به منظور اعمال یک قید احتمالاتی برای نزدیک نمودن توزیع مدل‌ها به توزیع واقعی استفاده می‌شود [۹]. همچنین می‌توان این دو مدل را در نقش دو عامل در یک بازی همکارانه قرار داده که با ارسال بازخوردهای موثر به یکدیگر سبب بهبود نتایج یکدیگر گردند [۸]. در این حالت، یادگیری با استفاده از روش‌های گرادیان سیاست انجام می‌شود. از همبستگی دوگان در سطح داده به روش‌های مختلفی ازجمله یادگیری بدون نظارت، یادگیری بانظارت، یادگیری نیمه‌نظارتی، یادگیری انتقالی و یادگیری تخاصمی می‌توان بهره‌گیری نمود که در ادامه به آن‌ها می‌پردازیم.

۱-۱-۳- یادگیری بدون نظارت دوگان

یادگیری بدون نظارت با استفاده از داده‌های بدون برچسب سعی بر یافتن الگوهای موجود در داده‌ها دارد. در یادگیری بدون نظارت دوگان آموزش وظایف اصلی و دوگان در تعامل با یکدیگر با استفاده از داده‌های بدون برچسب صورت می‌گیرد، هرچند ممکن است آموزش اولیه‌ی دو مدل نیاز به داده‌های برچسب‌دار داشته باشد. اولین بار ایده‌ی استفاده از یادگیری دوگان در ترجمه‌ی ماشینی عصبی به کار گرفته شد. یادگیری دوگان در ترجمه‌ی ماشینی [۸] سازوکاری[۲۱] را ارائه نمود که در آن از داده‌های تک‌زبانه[۲۲] (هم در زبان مبدا و هم در زبان مقصد) به‌صورتی کارا استفاده می‌شد به‌گونه‌ای که این داده‌ها نقشی مشابه پیکره‌ی دوزبانه‌ی موازی داشتند و به‌این‌ترتیب نیاز به پیکره‌ی دوزبانه‌ی موازی در آموزش مدل مترجم به مقدار قابل‌ملاحظه‌ای کاهش می‌یافت.

سازوکار یادگیری دوگان در ترجمه‌ی ماشینی را همان‌طور که در شکل ۱ مشاهده می‌کنید، می‌توان به‌صورت یک بازی دوعامله بیان نمود. فرض کنید دو عامل داریم که عامل اول تنها زبان فارسی و عامل دوم تنها زبان انگلیسی را درک می‌کند. عامل اول پیام x (به زبان فارسی) را از طریق یک کانال نویزی به عامل دوم می‌فرستد. این کانال پیام x را با استفاده از یک مدل مترجم به پیام y در زبان انگلیسی ترجمه می‌نماید. عامل دوم پیام y را دریافت نموده و با استفاده از مدل زبانی خود ارزیابی می‌کند که این پیام در زبان انگلیسی چقدر متداول است. سپس آن را از طریق کانال نویزی دیگر که پیام را با استفاده از یک مدل مترجم دیگر از زبان انگلیسی به فارسی ترجمه می‌کند، به عامل اول بازمی‌گرداند.

به‌این‌ترتیب، دو بازخورد جهت ارزیابی و بهبود مدل‌های مترجم دریافت می‌شود. بازخورد نخست از سازگاری پیام اصلی و پیام بازگشت داده شده به عامل اول با کمک معیار BLEU و بازخورد دوم از امتیازی که مدل زبانی عامل دوم به پیام ترجمه شده می‌دهد، به دست می‌آید. در این صورت با در نظر گرفتن دو پاداش برای این دو بازخورد می‌توان با روش‌های گرادیان سیاست، پارامترهای مدل‌های مترجم را به‌گونه‌ای به‌روزرسانی نمود که پاداش حاصل از این بازخوردها بیشینه شود. به همین شکل، در جهت عکس هم می‌توان مدل‌های مترجم را آموزش داد. این فرآیند ادامه می‌یابد تا هر دو مدل همگرا شوند و به دقت خوبی دست یابند. به‌این‌ترتیب یادگیری دوگان را می‌توان به‌گونه‌ای شبیه به یک بازی دو نفره دانست که در آن هر دو بازیکن سعی دارند در تعاملی همکارانه با یک‌دیگر، توانایی خود را ارتقا دهند.

اگرچه در این روش از داده‌های برچسب‌دار برای ایجاد یک مدل پایه‌ی قابل قبول استفاده شد، اما آموزش به روش یادگیری دوگان برای بهبود این مدل پایه و تنها با استفاده از داده‌های بدون برچسب صورت می‌گیرد. ازاین‌رو، این روش را در دسته‌ی یادگیر دوگان بدون نظارت قرار می‌دهند. درواقع، با این روش به‌گونه‌ای

اندازه‌ی مجموعه داده به‌صورت ضمنی افزایش یافته و در ادامه‌ی آموزش مدل‌های پایه با استفاده از آن علاوه بر بهبود عملکرد، مدل‌های پایدارتری به دست می‌آید.

۲-۱-۳- یادگیری بانظارت دوگان

در یادگیری بانظارت دوگان، همان طور که از نام آن برمی‌آید، تنها از داده‌های برچسب‌دار جهت آموزش دو مدل استفاده می‌شود. فرض کنید در این مسائل وظایف اصلی و دوگان را به ترتیب با توابع $f: X \to Y$ و $g: Y \to X$ و تابع هزینه‌ی آن‌ها را با ℓ_f و ℓ_g نمایش دهیم. جهت آموزش دو مدل با در نظر گرفتن ساختار دوگان از عبارت منظم‌ساز استفاده شده است [۹]. یک روش رایج برای طراحی f و g بیشینه‌سازی درست‌نمایی بر روی توزیع‌های شرطی $P(y|x;\ \theta_{xy})$ و $P(x|y;\ \theta_{yx})$ می‌باشد:

$$f(x;\theta_{xy}) \triangleq \underset{y' \in Y}{\operatorname{argmax}}\, P(y'|x;\theta_{xy}) \qquad (۱)$$

$$g(y;\theta_{yx}) \triangleq \underset{x' \in X}{\operatorname{argmax}}\, P(x'|y;\theta_{yx}) \qquad (۲)$$

که در آن θ_{xy} و θ_{yx} به ترتیب پارامترهای مدل‌های وظیفه‌ی اصلی و دوگان است. در حالت استاندارد یادگیری بانظارت، مدل f با استفاده از تابع هدف زیر آموزش داده می‌شود:

$$\min_{\theta_{xy}} \frac{1}{n} \sum_{i=1}^{n} \ell_f\left(f(x_i;\theta_{xy}), y_i\right) \qquad (۳)$$

و به همین ترتیب تابع هدف مدل g برای یادگیری به‌صورت زیر است:

$$\min_{\theta_{yx}} \frac{1}{n} \sum_{i=1}^{n} \ell_g\left(g(y_i;\theta_{yx}), x_i\right). \qquad (۴)$$

از طرفی با توجه به قانون بیز، اگر این دو مدل به خوبی و بی‌نقص آموزش ببینند، خواهیم داشت:

$$P(x)P(y|x;\ \theta_{xy}) = P(y)P(x|y;\ \theta_{yx}) = p(x,y) \quad \forall x, y. \qquad (۵)$$

برای برقراری رابطه‌ی (۵)، به تابع هدف مدل‌ها یک عبارت منظم‌ساز اضافه می‌شود تا توزیع‌های یادگرفته شده توسط مدل‌ها، به توزیع واقعی بیشتر نزدیک شود. عبارت منظم‌ساز را به‌صورت زیر تعریف می‌نماییم:

$$\ell_{duality}(x,y;\theta_{xy},\theta_{yx}) = \left(\log \hat{P}(x) + \log P(y|x;\ \theta_{xy}) - \log \hat{P}(y) - \log P(x|y;\ \theta_{yx})\right)^2 \qquad (۶)$$

که در آن $\hat{P}(x)$ و $\hat{P}(y)$ توزیع حاشیه‌ای تجربی مجموعه داده روی x و y می‌باشد. درنهایت تابع هدف یادگیری بانظارت دوگان برای وظیفه‌ی اصلی به‌صورت زیر به دست می‌آید:

$$L_f(\theta_{xy}) = \left(\frac{1}{m}\right) \sum_{i=1}^{m} \left[\ell_f\left(f(x_i), y_i;\ \theta_{xy}\right) + \lambda_{xy} \ell_{duality}(x_i, y_i;\ \theta_{xy}, \theta_{yx})\right] \qquad (۷)$$

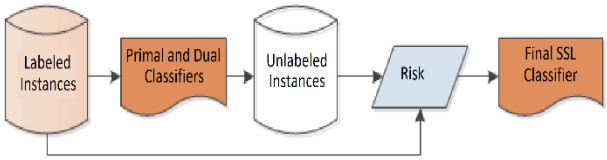
که در آن λ_{xy} ضریب مصالحه میان دو تابع هزینه ℓ_f و $\ell_{duality}$ می‌باشد. به‌صورت مشابه تابع هدف مدل g هم به دست می‌آید و سپس دو مدل به‌طور همزمان با روش نزول در امتداد گرادیان[۲۳] آموزش داده می‌شوند [۹]. در شکل ۲ الگوریتم یادگیری بانظارت دوگان را مشاهده می‌کنید. این روش در سه کاربرد ترجمه‌ی ماشینی، تحلیل تمایل و پردازش تصاویر پیاده‌سازی شد و در هر سه، نسبت به مدل‌های آموزش‌داده شده به‌صورت جداگانه بهبود قابل توجهی یافته بود [۹].

Input: Marginal distributions $\hat{P}(x_i)$ and $\hat{P}(y_i)$ for any $i \in [n]$; Lagrange parameters λ_{xy} and λ_{yx}; Optimizers Opt_1 and Opt_2;
repeat
Get a minibatch of m pairs $\{(x_j, y_j)\}_{j=1}^{m}$;
Calculate the gradient as follows:
$G_f = \nabla_{\theta_{xy}} (1/m) \sum_{j=1}^{m} [\ell_f(f(x_j), y_j;\ \theta_{xy}) + \lambda_{xy} \ell_{duality}(x_j, y_j;\ \theta_{xy}, \theta_{yx})]$
$G_g = \nabla_{\theta_{yx}} (1/m) \sum_{j=1}^{m} [\ell_g(g(y_j), x_j;\ \theta_{yx}) + \lambda_{yx} \ell_{duality}(x_j, y_j;\ \theta_{xy}, \theta_{yx})]$
Update the parameters of f and g:
$\theta_{xy} \leftarrow Opt_1(\theta_{xy}, G_f), \quad \theta_{yx} \leftarrow Opt_2(\theta_{yx}, G_g).$
until models converged

شکل ۲- الگوریتم یادگیری بانظارت دوگان [۹]

۳-۱-۳- یادگیری نیمه‌نظارتی دوگان

در یادگیری نیمه‌نظارتی آموزش با کمک مجموعه داده‌های برچسب‌دار و همچنین مجموعه داده‌های بدون برچسب صورت می‌گیرد. در بسیاری از کاربردهای دنیای واقعی، معمولا نمونه‌های دارای برچسب محدود بوده، در حالی که نمونه‌های بدون برچسب به مقدار کافی موجود است. به همین دلیل یادگیری نیمه‌نظارت توجه زیادی را به خود جلب کرده است، زیرا ابزاری مؤثر جهت بهره‌گیری از نمونه‌های بدون برچسب است. روش‌های سنتی یادگیری نیمه‌نظارتی فرض می‌کنند که تمامی نمونه‌ها در جهت بهبود عملکرد مفید هستند. در حالی که این طور نیست و در چندین پژوهش به‌طور نظری و تجربی نشان داده شده است که در برخی موارد نمونه‌های بدون برچسب می‌توانند سبب انحطاط عملکرد مدل شوند [۱۷]. بنابراین چگونگی استفاده ایمن[۲۴] از نمونه‌های بدون برچسب یک مسئله‌ی در حال ظهور و جالب در یادگیری نیمه‌نظارتی است. ازاین‌رو، روشی بر پایه‌ی یادگیری دوگان جهت یادگیری نیمه‌نظارتی امن[۲۵] به نام دالاس[۲۶] ارائه شده است که از یادگیری دوگان برای برآورد ایمنی یا ریسک[۲۷] نمونه‌های بدون برچسب بهره می‌برد.



شکل ۳- الگوریتم دالاس [۱۷]

ایده اصلی استفاده از یادگیری بانظارت برای تحلیل ریسک نمونه‌های بدون برچسب است. الگوریتم دالاس (شکل ۳) ابتدا از یک مدل اصلی به‌دست‌آمده توسط یادگیری دوگان برای طبقه‌بندی هر نمونه بدون برچسب و سپس از مدل دوگان برای بازسازی این نمونه‌ها با توجه به نتایج دسته‌بندی به‌دست‌آمده استفاده می کند. به‌این‌ترتیب، می‌توان ریسک را با تحلیل خطای بازسازی و دسته‌بندی نمونه‌های بدون برچسب واقعی و بازسازی‌شده‌ی آن اندازه‌گیری نمود. اگر خطای بازسازی کوچک و پیش‌بینی دسته‌بند برای هر دو نمونه برابر باشد، آن‌گاه احتمالا نمونه‌ی بدون برچسب ایمن است. در غیر این صورت، این نمونه ممکن است ریسک زیادی داشته باشد و خروجی آن باید به خروجی به‌دست‌آمده از یادگیری بانظارت نزدیک شود.

برای تحقق این هدف یک عبارت منظم‌ساز بر اساس ریسک به مدل یادگیری نیمه‌نظارتی اضافه می‌شود [۱۷]. به‌این‌ترتیب، خروجی‌های این الگوریتم مصالحه‌ای میان خروجی مدل بانظارت و مدل نیمه‌نظارتی است. برای بررسی کارآمدی این روش، مجموعه‌ای از آزمایش‌ها بر روی چندین مجموعه داده جهت مقایسه با روش‌های یادگیری بانظارت، نیمه‌نظارتی و نیمه‌نظارتی ایمن انجام داده شد. نتایج نشان می‌داد که دالاس می تواند به‌طور موثری ریسک نمونه‌های بدون برچسب را کاهش دهد و در هیچ مورد گزارش شده، دالاس عملکرد بدتر قابل‌ملاحظه‌ای نسبت به روش‌های یادگیری بانظارت و نیمه‌نظارتی ندارد [۱۷].

۳-۱-۴- یادگیری انتقالی دوگان

دانش به‌دست‌آمده از حل برخی مسائل می‌تواند به حل مسائل دیگر کمک نماید. یادگیری انتقالی یک روش یادگیری ماشینی است که در آن دانش به‌دست‌آمده از یک مسئله در حل مسئله‌ی دیگر به کار برده می‌شود. ونگ و همکارانش از راهبرد یادگیری انتقالی دوگان در مترجم ماشینی استفاده نمودند [۱۸]. به‌طور کلی، مترجم ماشینی قصد دارد تا نگاشتی از فضای زبانی X به فضای زبانی Y به دست آورد و این هدف با بیشینه نمودن درست‌نمایی توزیع $P(y|x;\theta)$ با مجموعه پارامتر θ، روی مجموعه داده‌ی $\{(x^{(i)},y^{(i)})\}_{i=1}^{n}$ که در آن $x^{(i)} \in X$ و $y^{(i)} \in Y$ است، محقق می‌شود. به سبب تعداد بسیار زیاد پارامترها در شبکه‌های ژرف مجموعه داده‌ی دوزبانه‌ی بزرگی لازم است تا بتوان به تخمین خوبی از پارامترهای توزیع مورد نظر دست یافت. اما گردآوری چنین مجموعه داده‌ای بسیار زمان‌بر و پرهزینه است. با انتقال دانش میان دو مدل وظیفه‌ی اصلی و دوگان به کمک یادگیری انتقالی دوگان می‌توان تا حدی از این مشکل کاست و قدرت تعمیم دو مدل را افزایش داد. ایده‌ی یادگیری انتقالی دوگان در ترجمه‌ی ماشینی برقرار نمودن رابطه‌ی زیر است:

$$P(y)=\sum_{x\in X}P(y|x;\theta)P(x)=\mathbb{E}_{x\sim P(x)}P(y|x;\theta)\cdot \qquad (۸)$$

درواقع انتظار می‌رود اگر توزیع به‌دست‌آمده توسط شبکه به توزیع واقعی نزدیک باشد، رابطه‌ی بالا که از قانون احتمال کل به دست می‌آید برقرار گردد [۱۸]. به‌این‌ترتیب، عبارت زیر را با ضریب مصالحه‌ی λ به تابع زیان استاندارد اضافه می‌گردد تا پارامترهای شبکه را بیشتر به سمت پارامترهای توزیع واقعی هدایت نماید:

$$S(\theta)=\left(\log\hat{P}(y)-\log\mathbb{E}_{x\sim\hat{P}(x)}P(y|x;\theta)\right)^2 \qquad (۹)$$

که در آن $\hat{P}(x)$ و $\hat{P}(y)$ توزیع تجربی به‌دست‌آمده با استفاده از مدل‌های زبانی آموزش دیده بر روی زبان مبدا و مقصد می‌باشند. از طرفی عبارت (۸) را می‌توان با نمونه‌برداری K نمونه از توزیع $P(x)$ به‌صورت زیر تخمین زد:

$$P(y)=\sum_{x\in X}P(y|x;\theta)P(x)=\mathbb{E}_{x\sim P(x)}P(y|x;\theta)$$
$$\approx\frac{1}{K}\sum_{i=1}^{K}P\left(y|x^{(i)};\theta\right),\quad x^{(i)}\sim P(x)\cdot \qquad (۱۰)$$

اما از آن‌جا که توزیع $P(y|x;\theta)$ برحسب x تنک است، این مقدار برای بیشتر نمونه‌ها نزدیک به صفر شده و تخمین $S(\theta)$ از این رابطه نمی‌تواند تاثیرگذار واقع شود. برای رفع این مشکل از نمونه برداری بر اساس اهمیت[۲۸] استفاده می‌شود که در آن از توزیع $P(x|y)$ بجای توزیع $P(x)$ نمونه‌برداری می‌شود. به‌این‌ترتیب با استفاده از یادگیری دوگان، توزیع $P(x|y)$ در مدل دوگان به دست آورده می‌شود و ازاین‌رو می‌توان به‌صورت کارا عبارت مورد نظر را محاسبه نمود [۱۸]. از آن جا که در این روش دانش به‌دست‌آمده در مدل دوگان به مدل اصلی و همچنین از مدل اصلی به مدل دوگان منتقل می‌شود، به آن یادگیری دوگان انتقالی می‌گویند.

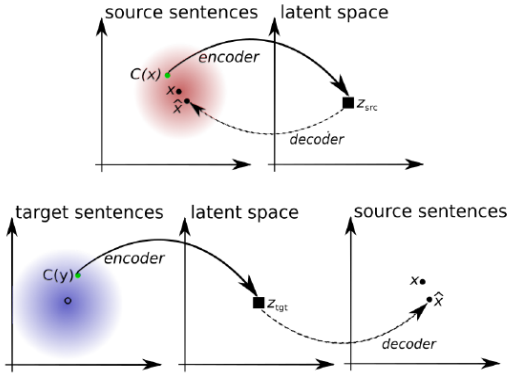
۳-۱-۵- یادگیری تخاصمی دوگان

از دیگر مباحث مورد توجه در سال‌های اخیر می‌توان به شبکه‌های تخاصمی مولد[۲۹] (GAN) اشاره نمود. این شبکه از دو شبکه‌ی ژرف شامل یک مولد[۳۰] و یک تمیزدهنده[۳۱] تشکیل شده است که در آن، این دو شبکه در قالب یک بازی مجموع صفر[۳۲] به‌طور همزمان آموزش داده می‌شوند. این بدین معنی است که دو شبکه در یک رقابت تخاصمی قرار دارند و بهبود یکی از این دو می‌تواند سبب شکست دیگری شود. هدف از آموزش این شبکه یافتن یک نقطه‌ی تعادل میان دو شبکه‌ی رقیب است. با استفاده ازاین‌روش نتایج فوق‌العاده‌ای در زمینه‌های مختلف ازجمله تولید متن [۱۹] و ترجمه‌ی تصویر به تصویر [۲۰، ۲۱] به‌دست‌آمده و از آن به عنوان تحولی در مدل‌های مولد یاد می‌شود. در ادامه به چند پژوهش که در آن از شبکه‌های تخاصمی مولد در قالب دوگان استفاده شده است، می‌پردازیم.

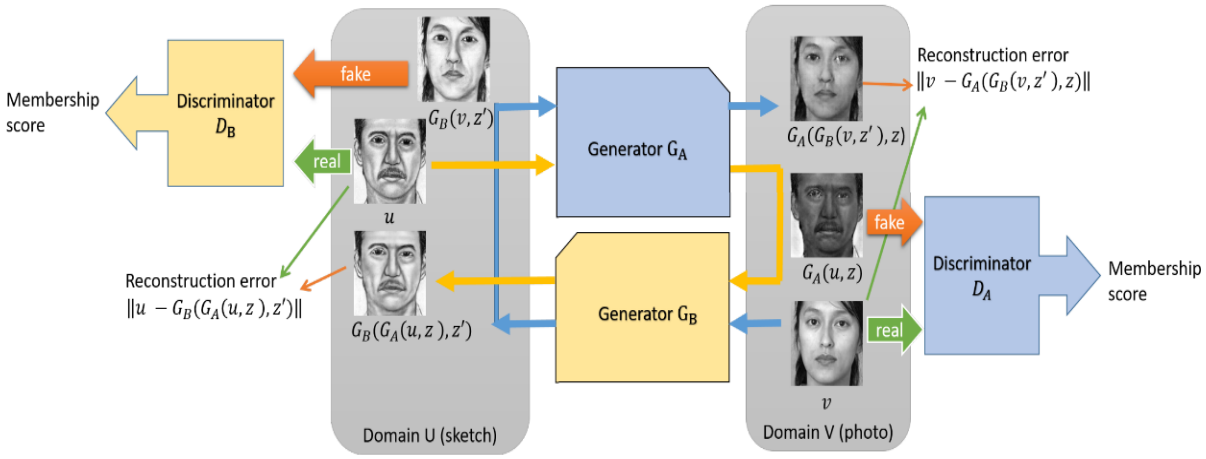
تا سال ۲۰۱۷ با ارائه‌ی روش‌های گوناگون در یادگیری ژرف پیشرفت قابل‌ملاحظه‌ای در مسئله‌ی ترجمه‌ی ماشینی صورت گرفت. اما این روش‌ها تنها در صورت وجود مجموعه داده‌های بزرگ دوزبانه کاربرد داشتند که با توجه به وجود زبان‌های متعدد، جمع‌آوری چنین مجموعه داده‌هایی برای هر جفت زبان امکان‌پذیر نیست. بدین ترتیب تلاش‌هایی جهت کاهش وابستگی به مجموعه داده‌های دوزبانه و بهره‌گیری از مجموعه‌های تک‌زبانه صورت گرفت [۲۲، ۲۳، ۲۴]. در پژوهشی توسط لمپل و همکاران، روشی برای حل مسئله ترجمه‌ی ماشینی بدون استفاده از هیچ مجموعه داده‌ی دوزبانه‌ای و تنها با استفاده از داده‌های تک‌زبانه ارائه گردید [۲۵].

ایده‌ی اصلی در این روش، ایجاد یک فضای نهفته[۳۳] میان دو دامنه (زبان) و یادگیری ترجمه با استفاده از بازسازی از این فضا براساس دو نکته بود. نخست این که مدل بتواند مطابق آن‌چه در شکل ۴ دیده می‌شود، یک جمله‌ی داده شده در یک زبان را از روی نسخه‌ی نویزی آن جمله و دوم، از روی ترجمه‌ی نویزی آن بازسازی نماید.

علاوه بر این دو نکته، با استفاده از یک عبارت منظم‌ساز تخاصمی، تلاش می‌شود تا بازنمایی نهفته‌ی جملات دو زبان مبدا و مقصد، توزیع یکسانی داشته باشند. به این صورت که یک تمیزدهنده که زبان را از روی بازنمایی نهفته‌ی داده شده تشخیص می‌دهد، به‌طور همزمان با مدل آموزش داده می‌شود و مدل تلاش می‌کند آن را گمراه نماید. این مراحل تکرار می‌شود تا مدل آموزش داده شده بهبود یابد. برای این که این روش کاملا بدون نظارت باشد، به عنوان مدل اولیه یک مترجم بدون نظارت ساده‌لوحانه از ترجمه‌ی کلمه به کلمه‌ای به کار برده شد که با استفاده از همان مجموعه داده‌ی تک‌زبانه به دست می‌آمد [۲۶].



شکل ۴- نمایش نحوه‌ی ترجمه از فضای بازنمایی نهفته در ترجمه‌ی ماشینی بدون نظارت [۲۵]



شکل ۵- معماری شبکه‌ی تخاصمی دوگان برای ترجمه تصویر به تصویر [۱۱]

در پردازش تصویر و بینایی ماشین مسائل متعددی ازجمله بخش‌بندی معنایی[۳۴] و انتقال سبک[۳۵] می‌تواند به عنوان یک مسئله ترجمه‌ی تصویر به تصویر مطرح شود که در آن بازنمایی تصویری یک شی به یک بازنمایی دیگر از آن تبدیل می‌گردد [۲۱]. اما با توجه به تفاوت‌های ذاتی موجود میان این مسائل، در بیشتر موارد به‌صورت جداگانه به آن‌ها پرداخته می‌شود. در سال‌های اخیر با استفاده از یادگیری ژرف، روش‌هایی عام‌منظوره[۳۶] برای حل این مسائل ارائه گردید. البته همه‌ی روش‌های ارائه‌شده به مجموعه داده‌های برچسب‌دار بزرگ نیاز داشتند و به‌صورت بانظارت آموزش داده می‌شدند. در پژوهشی توسط یی و همکاران در سال ۲۰۱۸ با استفاده از یادگیری دوگان در شبکه‌های تخاصمی، مسئله‌ی ترجمه‌ی تصویر به تصویر به‌صورت کاملا بدون نظارت انجام شد که در آن تنها از دو مجموعه تصویر از دو سبک مختلف استفاده می‌نمود [۱۱]. به وضوح، بزرگترین چالش در این مسئله، چگونگی تشخیص ترجمه‌ی درست تصاویر بود. این ایده بسیار شبیه به یادگیری دوگان در ترجمه‌ی ماشینی [۸] بود با این تفاوت که در این مسئله آموزش مدلی مشابه مدل زبانی در ترجمه‌ی دوگان جهت تشخیص مناسب بودن ترجمه بسیار سخت بود [۱۱]. به همین دلیل از ساختار شبکه‌های تخاصمی مولد استفاده شد.

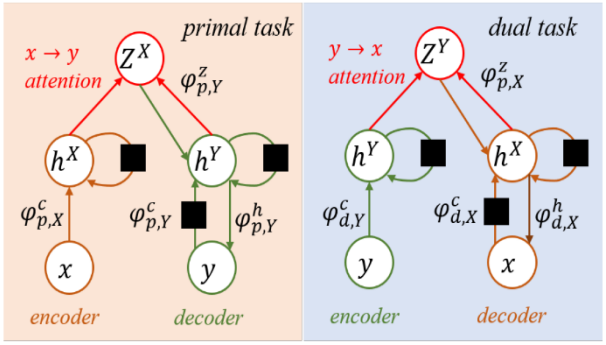
برای هر مجموعه از تصاویر از یک شبکه مولد تخاصمی استفاده شد و این دو شبکه در یک مسیر بسته همان‌طور در شکل ۵ مشاهده می‌شود، با هم در تعامل بوده و با استفاده از یادگیری دوگان آموزش داده می‌شدند. دو مجموعه تصویر بدون برچسب از دو دامنه‌ی X و Y را در نظر بگیرید. در این نوع یادگیری دوگان، وظیفه‌ی اصلی آموزش مدل مولد $G_A: X \to Y$ است که یک تصویر $u \in X$ را به یک تصویر $v \in Y$ نگاشت می‌دهد. به همین شکل، وظیفه‌ی دوگان آموزش مدل مولد $G_B: Y \to X$ می‌باشد. علاوه بر آن، وظیفه‌ی اصلی از یک تمیزدهنده D_A بهره می‌برد که وظیفه‌ی آن تشخیص تصاویر تولید شده توسط G_A از تصاویر واقعی در دامنه‌ی Y است. به همین ترتیب، وظیفه‌ی دوگان از تمیز دهنده‌ی D_B جهت تفکیک تصاویر تولید شده توسط G_B و تصاویر واقعی در دامنه‌ی X استفاده می‌نماید. درواقع، این تمیزدهنده‌ها نقشی همانند مدل زبانی در مترجم ماشینی دوگان [۸] دارد که با استفاده از آن می‌توان فهمید که تصاویر تولید شده توسط مولدها چقدر خوب در دامنه‌ی مورد نظر جای می‌گیرند [۱۱].

در این مدل نیز دو بازخورد داریم. یک بازخورد از تمیزدهنده که نشان می‌دهد تصویر ورودی چقدر خوب به دامنه‌ی مورد نظر ترجمه شده است و بازخورد دیگر که از خطای بازسازی تصویر تولید شده در یک گردش کامل توسط دو مولد (ترجمه تصویر به دامنه‌ی دیگر توسط مولد اول و سپس بازگشت به همان دامنه توسط مولد دوم) و تصویر واقعی به دست می‌آید. آموزش دو مدل با استفاده از این دو بازخورد تا رسیدن به نقطه‌ی تعادل ادامه یافته و درنهایت دو مدل قادر به ترجمه‌ی تصویر ورودی به دامنه‌ی دیگر خواهد بود.

۲-۳- یادگیری دوگان در سطح معماری مدل

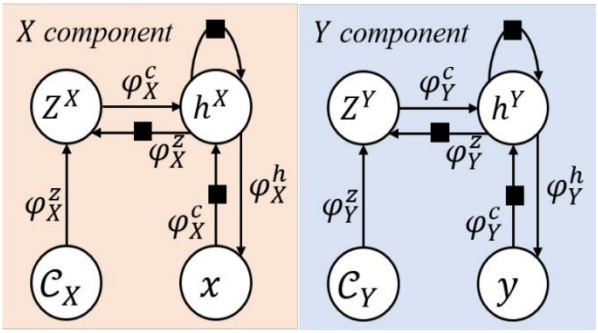
در بخش پیش، مروری کلی بر روش‌های یادگیری دوگان در سطح داده داشتیم و مشاهده نمودیم که به روش‌های متعددی می‌توان از دوگانگی موجود میان وظایف در سطح داده بهره برد. البته عمده‌ی تمرکز یادگیری دوگان در دوگانگی در سطح داده قرار دارد. در ادامه ایده‌ی دیگری در مبحث یادگیری دوگان که استفاده از دوگانگی در سطح معماری مدل است را بررسی خواهیم نمود.

در ادامه‌ی بهره‌گیری از همبستگی میان مسئله‌ها از دیدگاه‌های مختلف، در این روش هدف استفاده از دوگانگی در معماری مدل وظایف اصلی و دوگان است. بسیاری از مدل‌ها در مسائل گوناگون ازجمله پردازش زبان طبیعی با استفاده از یک ساختار کدگذار-کدگشا طراحی می‌شوند که در آن، ابتدا کدگذار[۳۷] ورودی را به یک بازنمایی در فضای نهفته می‌نگارد و سپس یک کدگشا[۳۸] خروجی مناسب را از روی بازنمایی ورودی در فضای نهفته به دست می‌آورد. برخی از این مسائل، مانند ترجمه‌ی ماشینی، دارای ساختاری متقارن بوده که در آن کدگذار و کدگشا عملی مشابه را در جهتی متفاوت انجام می‌دهند. در مقابل، برخی مسائل دیگر مانند تحلیل تمایل (تبدیل چندین کلمه‌ی ورودی به فضای نهان و سپس از فضای نهان به یک خروجی) و تولید تمایل، دارای ساختاری غیرمتقارن می‌باشند. با الهام از روش‌هایی مانند یادگیری چندوظیفه‌ای[۳۹] و اشتراک بازنمایی[۴۰] (که در آن لایه‌های پایینی مدل شبکه عصبی برای یادگیری وظایف مرتبط به اشتراک گذاشته می‌شوند و به‌این‌ترتیب به نتایجی بهتری می‌رسند [۲۷]) و با در نظر گرفتن ساختار دوگان مسائلی مانند

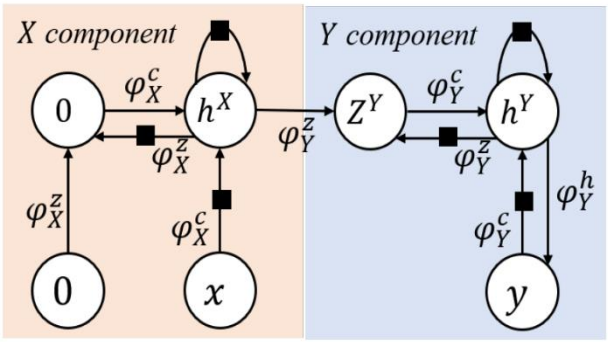


شکل ۶- در سمت چپ مدل وظیفه‌ی اصلی و در سمت راست مدل وظیفه‌ی دوگان قابل مشاهده است [۱۴].

ترجمه‌ی ماشینی و تحلیل و تولید تمایل، ایده‌ی به اشتراک‌گذاری پارامترهای مدل وظایف اصلی و دوگان به ذهن می‌رسد.



شکل ۷- در سمت چپ مدل وظیفه‌ی اصلی و در سمت راست مدل وظیفه‌ی دوگان قابل مشاهده است [۱۴].



شکل ۸- یادگیری دوگان در سطح معماری مدل [۱۴]

در مسائل متقارن می‌توان پارامترهای کدگذار در وظیفه‌ی اصلی و کدگشا در وظیفه‌ی دوگان و به‌طور مشابه، پارامترهای کدگذار در وظیفه‌ی دوگان و کدگشا در وظیفه‌ی اصلی را به اشتراک گذاشت. به‌این‌ترتیب، از دوگانگی موجود میان وظایف در سطح معماری بهره‌گیری می‌شود. با ایده‌ای مشابه در مسائل نامتقارن همانند تحلیل تمایل و تولید تمایل نیز پارامترهای کدگذار در وظیفه‌ی اصلی و کدگشا در وظیفه‌ی دوگان به اشتراک گذاشته شده و از دوگانگی در سطح معماری مدل بهره‌گیری می‌شود. با استفاده از این ایده، تعداد پارامترهای مدل به‌صورت کارا کاهش یافته و علاوه بر عمومیت[۴۱] بیشتر، پیچیدگی نمونه‌ای مدل کاهش می‌یابد.

مسئله‌ی ترجمه‌ی ماشینی میان زبان انگلیسی و فارسی را در نظر بگیرید. در شکل ۶ شمای معماری کدگذار-کدگشا مدل‌های وظایف اصلی و دوگان آن را مشاهده می‌کنید. در این شکل، φ پارامترهای مدل، h^X و h^Y به ترتیب بازنمایی نهان x و y، Z بازنمایی حاصل پس از اعمال سازوکار توجه بر روی بازنمایی‌های نهان h و هر مربع سیاه نشان‌دهنده‌ی یک واحد تاخیر زمانی است. جهت پیاده‌سازی یادگیری دوگان در سطح معماری مدل، دو جزء[۴۲] جداگانه (شکل ۷) برای هر زبان در نظر گرفته می‌شود. در سمت چپ شکل ۷، کدگذار زبان انگلیسی در وظیفه‌ی اصلی را مشاهده می‌کنید که همزمان کدگشای این زبان در وظیفه‌ی دوگان نیز می‌باشد. به همین ترتیب، کدگذار و کدگشای زبان فارسی در سمت راست این شکل قرار دارد. با توجه به تفاوت میان کدگذار و کدگشا (سازوکار توجه)، C_X که نشان‌دهنده‌ی اطلاعات محتوایی در بکارگیری توجه است به این دو جزء اضافه می‌شود. در شکل ۸ چگونگی ارتباط این دو مدل در هنگام آموزش وظیفه‌ی اصلی با روش یادگیری دوگان در سطح مدل را مشاهده می‌کنید.

در کدگذار زبان انگلیسی که نیازی به سازوکار توجه نیست، پارامترهای C_x و Z^X را برابر صفر قرار می‌دهیم و ورودی، با پارامتر φ_X^c به بازنمایی نهان h^X تبدیل می‌شود و این بازنمایی نهان به عنوان اطلاعات محتوایی در اختیار کدگشای زبان فارسی قرار می‌گیرید [۱۴]. دقیقا همین ماجرا برای ترجمه در وظیفه‌ی دوگان تکرار می‌شود. به‌این‌ترتیب در این دو مدل پارامترهای کدگذار در یک وظیفه و کدگشا در وظیفه‌ی دیگر به اشتراک گذاشته شده و تعداد پارامترهای مدل به شدت کاهش می‌یابد و امید می‌رود که قدرت تعمیم‌پذیری مدل افزایش یابد [۱۴].

۳-۳- استنتاج دوگان

تا به این‌جا در مورد دو روش برای استفاده از ساختار دوگان مسائل در مرحله‌ی آموزش صحبت نمودیم. نکته‌ی جالب این است که پس از آموزش دو مدل در مرحله‌ی استنتاج نیز می‌توان از هر دو مدل بهره برد. این روش که در پژوهشی با عنوان استنتاج دوگان ارائه گردید، یک چارچوب کاری عمومی برای بهره‌گیری از ساختار دوگان مسئله در هنگام استنتاج ارائه می‌نماید به‌گونه‌ای که بدون نیاز به آموزش دوباره، هر دو مدل وظیفه‌ی اصلی و دوگان در مرحله‌ی استنتاج هر یک از وظایف بکار گرفته می‌شود [۱۵]. شهود این استنتاج به این صورت است که y یک خروجی مناسب برای x در وظیفه‌ی اصلی می‌باشد، اگر و تنها اگر x نیز یک خروجی مناسب برای y در وظیفه‌ی دوگان باشد [۱۵].

بیش‌تر روش‌های استنتاجی که در یادگیری ماشین بکار گرفته می‌شود به‌صورت زیر عمل می‌کنند:

$$f(x) = \underset{y' \in Y}{\operatorname{argmin}}\, \ell_f(x, y');$$
$$g(y) = \underset{x' \in X}{\operatorname{argmin}}\, \ell_g(y, x'). \qquad (۱۱)$$

در استنتاج دوگان با دخالت هر دو مدل در مرحله‌ی استنتاج، عبارات (۱۱) به‌صورت زیر تغییر می‌کند:

$$f_{dual}(x) = \underset{y' \in Y}{\operatorname{argmin}}\{\alpha \ell_f(x, y') + (1-\alpha)\ell_g(y', x)\};$$
$$g_{dual}(y) = \underset{x' \in X}{\operatorname{argmin}}\{\beta \ell_g(y, x') + (1-\beta)\ell_f(x', y)\}; \qquad (۱۲)$$

که α و β پارامترهایی برای مصالحه بین دو تابع هزینه هستند و مقدار آن‌ها با کمک مجموعه داده‌ی اعتبارسنجی تعیین می‌شوند. بنابراین روابط (۱۲) حالت عمومی‌تری از عبارات (۱۱) هستند و می‌توانند به نتایج بهتری منجر شوند [۱۵]. برای نشان دادن اثر این روش استنتاج، از آن در سه مسئله‌ی ترجمه‌ی ماشینی، تحلیل تمایل و پردازش تصویر استفاده شده است و نتایج در هر سه مسئله بهبود مناسبی را نشان می‌دهند [۱۵].

۴- کاربردهای یادگیری دوگان ژرف

همان‌طور که تا به این‌جا شاهد بودیم، میان مسائل مختلف همبستگی و دوگانگی وجود دارد و می‌توان این مسائل را در قالب دوگان تعریف نمود. همچنین، دیدیم که چگونه با استفاده از یادگیری دوگان به روش‌های گوناگون می‌توان از این همبستگی بهره برد. در این بخش قصد داریم کاربردهای این روش‌ها در مسائل مختلف و نتایج به‌دست‌آمده در پژوهش‌های صورت گرفته را بررسی نماییم.

ترجمه‌ی ماشینی یکی از محبوب‌ترین مسائل در یادگیری دوگان می‌باشد به‌گونه‌ای که بیشترین تعداد پژوهش‌های این روش، بر روی مسئله ترجمه‌ی ماشینی صورت گرفته است. معمولا مترجم‌های ماشینی عصبی به‌صورت شبکه‌های عصبی کدگذار-کدگشا پیاده‌سازی می‌شوند و سعی دارند توزیع احتمال شرطی P(x|y) را تخمین بزنند که در آن x یک جمله در زبان مبدا و y یک جمله در زبان مقصد است. در این چارچوب، ابتدا جمله‌ی ورودی توسط کدگذار به یک فضای بازنمایی نهفته نگاشت داده شده و سپس ترجمه‌ی آن از روی این بازنمایی توسط کدگشا، معمولا به‌صورت کلمه به کلمه، تولید می‌شود. تابع هدف در آموزش استاندارد یک مدل

مترجم عصبی، بیشینه نمودن تابع درست‌نمایی روی مجموعه داده‌ی آموزشی است [۲۸]. ترجمه‌ی ماشینی یکی از مباحث داغ و پراهمیت در یادگیری ماشین و پردازش زبان طبیعی است و از آن جهت که در این مسئله هر دو وظیفه‌ی اصلی و دوگان متقارن بوده (هر دو وظیفه ترجمه‌ی ماشینی است) و اتلاف اطلاعات[۴۳] وجود ندارد، به خوبی در یادگیری دوگان جای می‌گیرد.

جدول ۱- مقایسه‌ی امتیاز BLEU مدل‌های مختلف مترجم در یادگیری دوگان

روش	En → Fr	Fr → En	En → De	De → En
یادگیری استاندارد (مدل پایه) [۲۸]	۲۹/۹۲	۲۷/۴۹	۱۶/۵۴	۲۰/۶۹
یادگیری دوگان بدون نظارت [۸]	۳۲/۰۶	۲۹/۷۸	۱۸/۴۹	۲۲/۱۴
یادگیری بانظارت دوگان [۹]	۳۱/۹۹	۲۸/۳۵	۱۷/۹۱	۲۰/۸۱
استنتاج دوگان [۱۵]	۳۰/۴۵	۲۷/۸۶	۱۷/۵۳	۲۱/۱۷
یادگیری بانظارت دوگان و استنتاج دوگان [۱۵]	**۳۲/۲۶**	**۳۰/۳۴**	**۱۸/۹۶**	**۲۲/۳۷**

اولین پیاده‌سازی از مترجم ماشینی با استفاده از یادگیری دوگان بدون نظارت در قالب یک بازی دو عامله انجام شد. با بکارگیری این روش و استفاده از تنها ده درصد از مجموعه‌ی داده‌ی برچسب‌دار برای یک شروع گرم[۴۴] و سپس آموزش بیشتر با استفاده از پیکره‌ی تک‌زبانه توانستند به بهبود قابل توجهی در ترجمه‌ی ماشینی از زبان فرانسوی به انگلیسی و بر عکس دست یابند، به‌گونه‌ای که امتیاز BLEU آن نزدیک به مدل‌های آموزش داده شده به‌صورت جداگانه بر روی تمام مجموعه داده بود [۸]. بنابراین یکی از مهم‌ترین کاربردهای این روش یادگیری کاهش نیاز به مجموعه داده‌ی برچسب‌دار بزرگ است که یکی از مهم بزرگترین چالش‌ها در یادگیری ژرف محسوب می‌شود. همچنین با استفاده از کل مجموعه داده‌ی برچسب‌دار توانست امتیاز BLEU مدل‌های مترجم در حالت دوگان را بیش از دو واحد نسبت به حالت استاندارد (آموزش به‌صورت جداگانه) بهبود دهد [۸]. در جدول ۱ مقایسه‌ی برخی از روش‌های به کار گرفته شده در ترجمه‌ی ماشینی را به همراه بخشی از نتایج مشاهده می‌کنید که در آن ترکیب یادگیری دوگان بانظارت و استنتاج دوگان بهترین نتیجه را کسب نموده است. همچنین، ترجمه‌ی ماشینی با استفاده از روش‌های دیگری هم انجام شده که در جدول ۲ نتایج آن و میزان بهبود نسبت به مدل پایه را مشاهده می‌نمایید. به وضوح مشاهده می‌شود که در بیشتر موارد استفاده از روش‌های گوناگون یادگیری دوگان به‌طور موثری سبب افزایش امتیاز BLEU شده است. لازم به ذکر است مدل پایه‌ی روش‌ها در جدول ۲ این دو روش متفاوت بوده و ازاین‌جهت قابل مقایسه نیستند.

روش یادگیری تخاصمی دوگان از دیگر روش‌هایی بود که با استفاده از آن یک مترجم ماشینی به‌صورت کاملا بدون نظارت و تنها با استفاده از مجموعه داده‌های تک‌زبانه آموزش داده شد و توانست سبب افزایش قابل ملاحظه‌ی امتیاز BLEU نسبت به مدل پایه (مترجم کلمه به کلمه) گردد [۲۵].

از دیگر مسائل مورد توجه در پردازش زبان طبیعی تحلیل تمایل است. در قالب یادگیری دوگان، تحلیل تمایل به عنوان وظیفه‌ی اصلی و تولید تمایل به عنوان وظیفه‌ی دوگان در نظر گرفته می‌شود. تحلیل یا استخراج تمایل، مطالعه‌ی محاسباتی نظرات، تمایل، احساسات و نگرش‌های مردم نسبت به محصولات، خدمات، سازمان ها، افراد، مسائل، حوادث، موضوعات و ویژگی های آنها است [۲۹]. بنابراین در این مسئله وظیفه‌ی اصلی دسته‌بندی تمایل متن یا جمله‌ی ورودی و وظیفه‌ی دوگان، تولید جمله یا متن با حفظ یک تمایل مشخص است.

جدول ۲- امتیاز BLEU به‌دست‌آمده از روش‌های مختلف در ترجمه‌ی ماشینی به همراه میزان بهبود نسبت به مدل پایه

روش	En → Fr	Fr → En	En → De	De → En
یادگیری دوگان انتقالی [۱۸]	۳۲/۸۵ (+۲/۹۳)	–	–	۳۲/۳۵ (+۱/۳۶)
یادگیری دوگان سطح مدل [۱۴]	–	–	۲۸/۶۴ (+۰/۹)	۳۴/۷۱ (+۱/۸۵)

مسئله‌ی تحلیل تمایل و تولید تمایل در قالب دوگان با استفاده از سه روش یادگیری بانظارت دوگان، یادگیری دوگان در سطح مدل و استنتاج دوگان بر روی مجموعه داده‌ی IMDB [۳۰] انجام شد که نتایج به‌دست‌آمده در جدول ۳ قابل مشاهده و مقایسه است. همان طور که مشاهده می‌شود یادگیری دوگان توانسته است همواره نسبت به مدل پایه بهبود قابل ملاحظه به دست آورد. یک مسئله‌ی جالبی که به آن پرداخته نشده است و نیاز به بررسی دارد ترکیب این سه روش برای حل مسئله‌ی تحلیل تمایل و تولید تمایل است. با این کار می‌توان دید یادگیری دوگان حداکثر چه مقدار سبب بهبود نتایج می‌شود.

جدول ۳- نتایج روش‌های مختلف برای مسئله‌ی تحلیل تمایل و تولید تمایل. مقدار سرگشتگی برای روش استنتاج دوگان گزارش نشده است.

روش	خطای آزمون (٪)	سرگشتگی
مدل پایه	۱۰/۱۰	۵۹/۱۹
یادگیری بانظارت دوگان	۹/۲۰	۵۸/۷۸
استنتاج دوگان	۸/۳۱	–
یادگیری دوگان در سطح مدل	۷/۴۱	۵۵/۵۹

با استفاده از روش یادگیری بانظارت دوگان در پژوهشی توسط تانگ و همکاران به حل مسائل پاسخ‌گویی به پرسش (QA) و تولید پرسش (QG)، به ترتیب به‌عنوان وظیفه‌ی اصلی و وظیفه‌ی دوگان، در قالب یادگیری دوگان پرداخته شد [۱۲]. به‌طور دقیق‌تر، وظیفه‌ی اصلی، مسئله‌ی انتخاب پاسخ [۳۱] بود که در آن پاسخ از میان چندین جمله‌ی ورودی انتخاب می‌شود. همچنین، وظیفه‌ی دوگان تولید پرسش برای جمله‌ی ورودی بود. به‌این‌ترتیب ورودی و خروجی این دو وظیفه (تقریباً) برعکس یکدیگر بوده و می‌توانند با یادگیری دوگان آموزش داده شوند [۱۲]. این روش بر روی سه مجموعه داده‌ی MARCO [۳۲]، SQUAD [۳۳] و WikiQA [۳۱] آموزش داده شد و نتایج در هر سه مورد نسبت به مدل پایه بهبود یافت. اما در پژوهشی دیگر در سال ۲۰۱۸ روشی نسبتاً متفاوت، با استفاده از یادگیری بانظارت دوگان (استفاده از یک تابع زیان دوگان) و همچنین اشتراک جزئی پارامترهای دو مدل (ایده‌ای مشابه یادگیری سطح مدل) در مسئله‌ی پاسخگویی به پرسش و تولید پرسش به کار گرفته شد. برخلاف پژوهش قبلی [۱۲]، در این پژوهش هر دو وظیفه به شکل تولید توالی در نظر گرفته شده و از معماری یکسان برای هر

دو استفاده شد [۳۴]. این معماری به‌صورت سلسله‌ای شامل نهفته‌سازی واژگان، کدگذاری، سازوکار توجه و درنهایت کدگشایی بود. در هرکدام از این مراحل، تمام یا بخشی از پارامترها بین مدل وظیفه‌ی اصلی و وظیفه‌ی دوگان به اشتراک گذاشته می‌شد. به‌این‌ترتیب علاوه بر انتقال دانش میان دو مدل، تعداد کل پارامترها کاهش می‌یافت و درنتیجه قدرت تعمیم دو مدل بیشتر می‌شد. این روش نیز بر روی سه مجموعه داده‌ی MARCO [۳۲]، SQUAD [۳۳] و WikiQA [۳۱] توانست در کل به مقدار قابل‌توجهی نسبت به مدل پایه در معیارهای گوناگون ازجمله BLEU، Rouge-L و Meteor بهبود ایجاد نماید [۳۴].

پاسخ به سؤالات بصری از دیگر مسائل مهم است که در آن به بررسی ارتباط میان پرسش و تصویر پرداخته می‌شود. وظیفه‌ی دوگان آن در چارچوب دوگان، تولید پرسش بصری است که در آن یک پرسش مرتبط با تصویر توسط مدل تولید می‌گردد. در پژوهشی توسط لی و همکاران با روشی از ترکیب روش یادگیری بانظارت دوگان و یادگیری سطح مدل توانست بر روی مجموعه داده‌ی CLEVR [۳۵] و VQA2 [۳۶]، مدل پیشرو قبلی را به ترتیب ۱/۳۳ و ۰/۸۸ درصد بهبود دهد [۳۷].

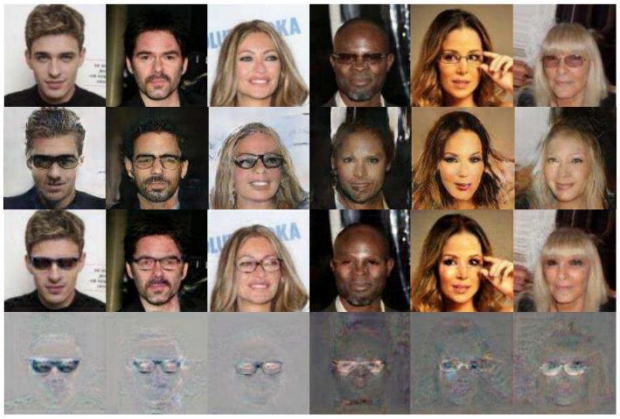
بسیاری از مسائل پردازش تصویر می‌تواند در قالب یک مسئله ترجمه‌ی تصویر

شکل ۱۰- ترجمه‌ی تصویر به تصویر با استفاده از روش تخاصمی دوگان [۱۱]. ستون سمت چپ، تصاویر ورودی، ستون میانی، خروجی شبکه‌ی تخاصمی دوگان و ستون سمت راست، خروجی شبکه‌ی تخاصمی استاندارد است.

به تصویر مطرح شود که در آن بازنمایی تصویری یک شی به یک بازنمایی دیگری تبدیل می‌گردد [۲۱]. ازآنجاکه ترجمه‌ی تصویر به تصویر می‌تواند شامل ترجمه‌ی تصاویر از دامنه‌های گوناگون و همچنین مسائل مختلفی ازجمله بخش‌بندی معنایی و انتقال سبک باشد، معمولاً به این مسائل به‌صورت جداگانه پرداخته می‌شود. البته دیدیم که با استفاده از یادگیری تخاصمی دوگان می‌توان به این مسائل به‌صورت عمومی پرداخت و همچنین این روش کاملاً به‌صورت بدون نظارت انجام می‌شود. مجموعه‌ای از تصاویر ایجاد شده با استفاده از این روش را در مقایسه با روش تخاصمی استاندارد در شکل ۱۰ مشاهده می‌نمایید. تغییر سن صورت از دیگر مسائلی است که با استفاده از روش تخاصمی دوگان شرطی[۴۵] انجام شده است که با افزودن یک شرط (سن صورت) به مدلی مشابه روش تخاصمی دوگان به دست می‌آید [۳۸].

با ایده‌ای کاملاً مشابه روش تخاصمی دوگان پژوهش دیگری در تغییر ویژگی‌های صورت انجام شد [۳۹]. برای نمونه، اضافه نمودن یا برداشتن عینک آفتابی از صورت اشخاص. تفاوت عمده‌ی این روش این بود که در آن تنها از یک تمیزدهنده و جهت دسته‌بندی به سه دسته‌ی تصویر تولید شده، تصویر واقعی با ویژگی مثبت (وجود ویژگی) و تصویر واقعی با ویژگی منفی (عدم وجود ویژگی) استفاده شد. همچنین در این مدل دو مولد برای تولید تصویر دارای یک ویژگی و تولید تصویر بدون یک ویژگی وجود دارد. با استفاده از این روش، یک تصویر که نشان‌دهنده‌ی تفاوت تصویر خروجی از تصویر ورودی بود، توسط شبکه‌ی مولد ایجاد شده و با تصویر ورودی جمع می‌گردید تا تصویر نهایی به دست آید. نمونه‌هایی از نتایج به‌دست‌آمده در این پژوهش را در شکل ۹ مشاهده می‌نمایید.

ازآنجاکه این‌گونه ترجمه‌ی تصویر به تصویر دارای تنوع کافی نبود و کنترل کافی برای ترجمه‌ی تصویر به دامنه‌ی مقصد وجود نداشت، در پژوهش دیگری با

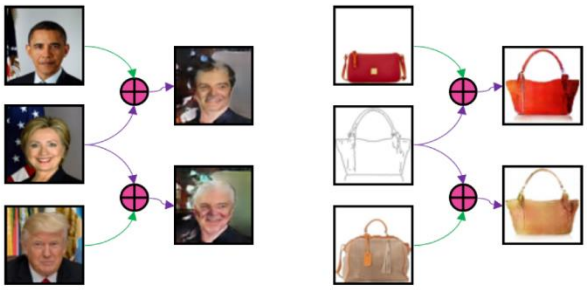


شکل ۹- تغییر ویژگی‌های صورت با استفاده از یادگیری تخاصمی دوگان [۳۹]

عنوان ترجمه‌ی شرطی تصویر به تصویر این مهم اعمال گردید [۱۰]. دو دامنه‌ی تصویر X و Y را در نظر بگیرید. فرض می‌شود هر تصویر $x \in X$ به‌صورت $x = x^i \oplus x^s$ نمایش داده می‌شود که در آن x^i و x^s به ترتیب ویژگی‌های مستقل از دامنه و ویژگی‌های وابسته به دامنه است و $\oplus$ عملگری است که با اعمال بر روی این دو مجموعه ویژگی، تصویر کامل x را ایجاد می‌نماید. به‌این‌ترتیب، ترجمه‌ی شرطی تصویر به تصویر از دامنه‌ی X به Y به‌صورت زیر تعریف می‌شود:

$$x_{AB} = G_{A\to B}(x, y) = x^i \oplus y^s \qquad (۱۳)$$

که در آن x^i ویژگی‌های مستقل از دامنه‌ی تصویر ورودی و y^s ویژگی‌های وابسته به دامنه‌ی تصویر شرط است. درواقع در این نوع ترجمه، همان‌طور که در شکل ۱۱ نیز مشاهده می‌شود، سبک تصویر خروجی با استفاده از یک تصویر دیگر تعیین می‌گردد. در این پژوهش، برای تولید تصاویر از ساختار کدگذار-کدگشا استفاده شد که در آن ابتدا دو کدگذار جداگانه تصویر ورودی و تصویر شرط را به فضای ویژگی‌های مستقل از دامنه و وابسته به دامنه برده و سپس با جابجا نمودن دو بردار ویژگی‌های وابسته به دامنه‌ی دو تصویر، بردارهای به‌دست‌آمده را به یک کدگشا می‌دهد [۱۰]. سپس با استفاده از دو تمیزدهنده‌ی جداگانه برای تصاویر هر دامنه تصاویر تولید شده را با تصاویر واقعی از آن دامنه مقایسه می‌نمود. به‌این‌ترتیب، مولدها را مجبور می‌سازد تا تصاویری در دامنه‌ی مورد نظر ایجاد نمایند. درواقع این



شکل ۱۱- ترجمه شرطی تصویر به تصویر [10]. سمت چپ، تبدیل تصویر زن به مرد و سمت راست، تبدیل تصویر حاشیه به کیف دستی است.

روش حالت تعمیم‌یافته روش تخاصمی دوگانی است که توسط یی و همکارانش در [۱۱] ارائه گردید و پیش‌تر در بخش ۳-۱ بررسی نمودیم که با حذف ویژگی‌های وابسته به دامنه به همان مدل قابل‌تبدیل است. در مقایسه با روش قبلی که تنها خطای بازسازی تصویر را در نظر می‌گرفت در این روش خطای بازسازی بردار ویژگی‌های مستقل از دامنه و وابسته به دامنه نیز به دست می‌آید و ازاین‌جهت انتظار می‌رود این روش تصاویر بهتری تولید نماید [۱۰].

تبدیل متن به گفتار و تشخیص گفتار دو وظیفه‌ی دوگان در حوزه‌ی پردازش گفتار می‌باشند که به لطف یادگیری ژرف و داده‌های زیاد متنی و گفتاری مناسب، پیشرفت‌های چشمگیری در سال‌های اخیر داشته‌اند. بااین‌حال، همچنان عدم وجود داده‌های برچسب‌دار در بسیاری از زبان‌ها یک ضعف بزرگ در حل این مسائل به شمار می‌رود. این دو مسئله از مسائل رایج دنباله به دنباله بوده و معمولاً با استفاده از یک چارچوب کدگذار-کدگشا در کنار یک سازوکار توجه پیاده‌سازی می‌گردند. برای حل این مسائل در قالب دوگان در پژوهشی با استفاده از یادگیری تقریباً بدون نظارت به حل این مسئله پرداخته شده است که در آن عمده‌ی آموزش بر روی مجموعه داده‌ی بدون برچسب صورت گرفته و تنها از چند صد نمونه‌ی برچسب‌دار استفاده می‌شد [۱۳]. به این صورت که بخشی از تابع زیان با استفاده از داده‌های برچسب‌دار اندک و بخش دیگری که مربوط به یادگیری دوگان است، تنها با استفاده از مجموعه‌ای از متن‌ها و گفتارها به دست می‌آمد. ابتدا مجموعه‌ای از گفتارها با استفاده از مدل تشخیص گفتار، به متن معادل تبدیل گشته و سپس این متن و گفتار همانند یک مجموعه داده‌ی برچسب‌دار برای آموزش مدل متن به گفتار استفاده می‌شد. به همین ترتیب در جهت عکس از مجموعه داده‌ی به‌دست‌آمده توسط مدل گفتار به متن برای آموزش مدل تشخیص گفتار استفاده شده است. این دو مدل با استفاده از ساختار خودکدگذار بر اساس ترنسفرمر[۴۶] [۴۰] پیاده‌سازی شدند. دو مدل متن به گفتار و تشخیص گفتار با استفاده از این روش و تنها ۲۰۰ زوج نمونه‌ی متن و گفتار از مجموعه داده‌ی LJSpeech [۴۱] و مقدار زیادی متن و گفتار بدون‌برچسب آموزش داده شد و امتیاز MOS[۴۷] مدل متن به گفتار به مقدار ۲/۶۸ و مقدار PER[۴۸] مدل تشخیص گفتار به ۱۱/۷ درصد رسید [۱۳].

همان‌گونه که دیدیم مسائل زیادی را می‌توان به سادگی در قالب یادگیری دوگان طراحی نمود. سپس این مسائل را با استفاده از روش‌های مختلفی که به آن‌ها اشاره شد حل نموده و نتایج را به مقدار قابل‌ملاحظه‌ای بهبود داد.

۵- جمع‌بندی و نتیجه‌گیری

در این مقاله دیدیم که بسیاری از مسائل در قالب دوگان ظاهر می‌شوند و می‌توان از این وابستگی در سطوح مختلفی بهره‌برداری نمود. پس مروری ساختار یافته بر روش‌های گوناگون یادگیری دوگان در سطوح مختلف داشتیم و سپس کاربردهای هر یک را در مسائل مختلف به همراه نتایج به‌دست‌آمده بررسی و مقایسه نمودیم. همچنین مشاهده نمودیم که یادگیری دوگان با کاهش تعداد پارامترهای دو مدل یا افزایش ضمنی اندازه‌ی مجموعه داده می‌تواند به مدل‌های پایدارتر با دقت بهتری دست یابد. به‌طورکلی، یادگیری دوگان یک روش عمومی است که به‌راحتی می‌تواند بر مدل‌های پیشرو کنونی اعمال شود.

تاکنون یادگیری دوگان در کاربردهای بسیاری پیاده‌سازی گردیده و بهبود نتایج به‌صورت تجربی دیده شده است. اما یکی از مهم‌ترین کارهایی که در پژوهش‌های آینده به آن می‌توان پرداخت بررسی نظری یادگیری دوگان است. یکی از مهم‌ترین سؤالات در این زمینه این است که یادگیری دوگان در چه مواقعی و تا چه حدی می‌تواند سبب بهبود نتایج شود. همچنین، بررسی تعامل روش‌های یادگیری دوگان در سطوح مختلف و تأثیر آن‌ها بر یکدیگر می‌تواند به درک هرچه بیشتر آن کمک نماید. درنهایت، پیاده‌سازی یک چارچوب برنامه‌نویسی متن‌باز برای یادگیری دوگان می‌تواند به‌سرعت پیشرفت و پژوهش‌های بیشتر در این زمینه کمک چشمگیری نماید.

مراجع

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no.7553, pp. 436, 2015.

[2] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, L. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, J. Dean, "Google's neural machine translation system: Bridging the gap between human and machine translation," arXiv preprint arXiv:1609.08144, 2016.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.

[4] A. Oord, N. Kalchbrenner, and K. Kavukcuoglu. "Pixel recurrent neural networks," arXiv preprint arXiv:1601.06759, 2016.

[5] D. Amodei, R. Anubhai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, J. Chen, M. Chrzanowski, A. Coates, G. Diamos, and E. Elsen, "End to end speech recognition in English and Mandarin," 2016.

[6] A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," In Proceedings of the IEEE international conference on acoustics, speech and signal processing, pp. 6645-6649, 2013.

[7] A. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," arXiv preprint arXiv:1609.03499, 2016.

[8] D. He, Y. Xia, T. Qin, L. Wang, N. Yu, T. Liu, and W. Ma, "Dual learning for machine translation," In Proceedings of the Advances in Neural Information Processing Systems, pp. 820-828. 2016.

[9] Y. Xia, T. Qin, W. Chen, J. Bian, N. Yu, and T. Liu, "Dual supervised learning," In Proceedings of the 34th International Conference on Machine Learning, vol. 70, pp. 3789-3798, 2017.

[10] J. Lin, Y. Xia, T. Qin, Z. Chen, and T. Liu, "Conditional image-to-image translation," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5524-5532, 2018.

[11] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," In Proceedings of the IEEE international conference on computer vision, pp. 2849-2857, 2017.

[12] D. Tang, N. Duan, T. Qin, Z. Yan, and M. Zhou, "Question answering and question generation as dual tasks," arXiv preprint arXiv:1706.02027, 2017.

[13] Y. Ren, X. Tan, T. Qin, S. Zhao, Z. Zhao, and T. Liu, "Almost Unsupervised Text to Speech and Automatic Speech Recognition," arXiv preprint arXiv:1905.06791, 2019.

[14] Y. Xia, X. Tan, F. Tian, T. Qin, N. Yu, and T. Liu, "Model-level dual learning," In Proceedings of the International Conference on Machine Learning, pp. 5383-5392, 2018.

[15] Y. Xia, J. Bian, T. Qin, N. Yu, and T. Liu, "Dual Inference for Machine Learning," In Proceedings of the International Joint Conferences on Artificial Intelligence, pp. 3112-3118, 2017.

[16] I. Goodfellow, Y. Bengio, and A. Courville, "Deep learning," MIT press, 2016.

[17] H. Gan, Z. Li, Y. Fan, and Z. Luo, "Dual learning-based safe semi-supervised learning," IEEE Access, vol. 6, pp. 2615-2621, 2017.

[18] Y. Wang, Y. Xia, L. Zhao, J. Bian, T. Qin, G. Liu, and T. Liu, "Dual transfer learning for neural machine translation with marginal distribution regularization," In Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 2018.

[19] Y. Zhang, Z. Gan, and L. Carin, "Generating text via adversarial training," In Proceedings of the Advances in Neural Information

Processing Systems workshop on Adversarial Training, vol. 21, 2016.

[20] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," In Proceedings of the IEEE international conference on computer vision, pp. 2223-2232, 2017.

[21] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125-1134, 2017.

[22] A. Irvine, and C. Callison-Burch, "Combining bilingual and comparable corpora for low resource machine translation," In Proceedings of the eighth workshop on statistical machine translation, pp. 262-270, 2013.

[23] A. Irvine, and C. Callison-Burch, "End-to-end statistical machine translation with zero or small parallel texts," Natural Language Engineering, vol. 22, no. 4, pp. 517-548, 2016.

[24] H. Zheng, Y. Cheng, and Y. Liu, "Maximum Expected Likelihood Estimation for Zero-resource Neural Machine Translation," In Proceedings of the International Joint Conferences on Artificial Intelligence, pp. 4251-4257, 2017.

[25] G. Lample, A. Conneau, L. Denoyer, and M. Ranzato, "Unsupervised machine translation using monolingual corpora only," arXiv preprint arXiv:1711.00043, 2017.

[26] A. Conneau, G. Lample, M. Ranzato, L. Denoyer, and H. Jégou, "Word translation without parallel data," arXiv preprint arXiv:1710.04087, 2017.

[27] D. Dong, H. Wu, W. He, D. Yu, and H. Wang, "Multi-task learning for multiple language translation." In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, vol. 1, pp. 1723-1732. 2015.

[28] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv:1409.0473, 2014.

[29] B. Liu, "Sentiment analysis: Mining opinions, sentiments, and emotions," Cambridge University Press, 2015.

[30] IMDB dataset, http://ai.stanford.edu/amaas/data/sentiment/, accessed in October 2011.

[31] Y. Yang, W. Yih, and C. Meek, "Wikiqa: A challenge dataset for open-domain question answering," In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 2013-2018, 2015.

[32] T. Nguyen, M. Rosenberg, X. Song, J. Gao, S. Tiwary, R. Majumder, and L. Deng, "MS MARCO: A Human-Generated MAchine Reading COmprehension Dataset," 2016.

[33] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "Squad: 100,000+ questions for machine comprehension of text," arXiv preprint arXiv:1606.05250, 2016.

[34] H. Xiao, F. Wang, J. Yan and J. Zheng, "Dual ask-answer network for machine reading comprehension," arXiv preprint arXiv:1809.01997, 2018.

[35] J. Johnson, B. Hariharan, L. Maaten, L. Fei-Fei, C. L. Zitnick, and R. Girshick, "Clevr: A diagnostic dataset for compositional language and elementary visual reasoning," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2901-2910, 2017.

[36] Y. Goyal, T. Khot, D. Summers-Stay, D. Batra, and D. Parikh, "Making the V in VQA matter: Elevating the role of image understanding in Visual Question Answering," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6904-6913, 2017.

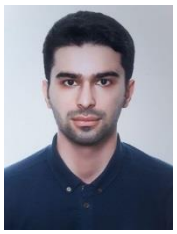
[37] Y. Li, N. Duan, B. Zhou, X. Chu, W. Ouyang, X. Wang, and M. Zhou, "Visual question generation as dual task of visual question answering," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6116-6124, 2018.

[38] J. Song, J. Zhang, L. Gao, X. Liu, and H. T. Shen, "Dual Conditional GANs for Face Aging and Rejuvenation," In Proceedings of the International Joint Conferences on Artificial Intelligence, pp. 899-905, 2018.

[39] W. Shen, and R. Liu, "Learning residual images for face attribute manipulation," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4030-4038, 2017.

[40] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. "Attention is all you need." In Proceedings of the Advances in neural information processing systems, pp. 5998-6008. 2017.

[41] K. Ito, The LJspeech dataset. https://keithito.com/LJ-Speech-Dataset/, accessed in November 2017.



علی‌اکبر خوش‌ویشکائی مدرک کارشناسی خود را در سال ۱۳۹۶ از دانشگاه علم و صنعت ایران در رشته‌ی مهندسی کامپیوتر دریافت نمود. او در سال ۱۳۹۹ در مقطع کارشناسی ارشد در گرایش هوش مصنوعی از دانشگاه صنعتی شریف فارغ‌التحصیل شد. زمینه‌های پژوهشی مورد علاقه‌ی او یادگیری دوگان عمیق، مدل‌های مولد عمیق و یادگیری عمیق بیزی است.

آدرس پست الکترونیکی ایشان عبارت است از:

khoshvishkaie@ce.sharif.edu

حمید بیگی تحصیلات خود را در مقاطع کارشناسی و کارشناسی ارشد رشته‌ی مهندسی کامپیوتر دانشگاه شیراز و سپس مقطع دکتری را در همان رشته از دانشگاه امیرکبیر به اتمام رساند. او هم اکنون به عنوان عضو هیئت علمی دانشکده مهندسی کامپیوتر دانشگاه صنعتی شریف مشغول فعالیت است. زمینه‌های پژوهشی مورد علاقه‌ی ایشان یادگیری ماشین در مقیاس بزرگ، داده‌کاوی و تئوری یادگیری ماشین است.

آدرس پست الکترونیکی ایشان عبارت است از:

beigy@sharif.edu

[1] Backpropagation
[2] Dual
[3] Dual learning
[4] Primal task
[5] Dual task
[6] Closed-loop learning
[7] Policy gradient methods
[8] Regularization term
[9] Model-level duality
[10] Inference
[11] Adversarial
[12] Artificial neuron
[13] Sigmoid
[14] Universal Approximator
[15] Sample complexity
[16] Graphical Processing Units
[17] Iterations
[18] Dual form
[19] Generalization
[20] Reconstruction error
[21] Mechanism
[22] Monolingual data
[23] Gradient descent
[24] Safe

[25] Safe semi-supervised learning
[26] DALLAS
[27] Risk
[28] Importance sampling
[29] Generative Adversarial Networks
[30] Generator
[31] Discriminator
[32] Zero-sum
[33] Hidden space
[34] Semantic segmentation
[35] Style transfer
[36] General purpose
[37] Encoder
[38] Decoder
[39] Multi-task learning
[40] Shared representation
[41] Generalization
[42] component
[43] Information loss
[44] Warm start
[45] Dual conditional GAN
[46] Transformer
[47] Mean Opinion Score
[48] Position-independent word error rate

The CSI Journal on Computing Science and Information Technology

The CSI Journal on
Computing Science and
Information Technology
Vol. 17, No.2, 2020

Deep Dual Learning and its Applications

Ali Akbar Khoshvishkaie, Hamid Beigy

Department of Computer Engineering, Sharif University of Technology, Tehran, Iran

Abstract

Deep learning, which has achieved the state-of-the-art results in many AI tasks, intensely suffers from the fact that its performance heavily depends on the scale of labeled data. In various real-world applications, labeled instances are generally limited and expensive to collect, while there are lots of unlabeled ones, the amount of which is often sufficient. Therefore, tools for effectively exploiting the unlabeled instances have attracted much attention. Apart from that, many AI tasks emerge in dual form, e.g., English-to-Persian translation vs. Persian-to-English translation, and image classification vs. image generation. Recently, several methods have been proposed to utilize the correlation between dual tasks. In this paper, we are going to review Dual Learning Methods, which aim to effectively employ the duality between two dual tasks in the training and/or inference. Dual learning can be divided into three different levels, namely data-level, model-level, and inference-level dualities. In this paper, we will look at different methods to use these ideas and their success in different applications. We will also demonstrate how Dual Learning effectively reduces demanding on labeled data.

Keywords: Deep Learning, Dual Learning, Dual Tasks.