برچسب‌گذاری تصاویر بدون نمونه آموزشی با کمک شبکه‌های عصبی بازگشتی

مهرداد باقری[1*] ، دکتر محرم منصوری‌زاده[2] ، دکتر میرحسین دزفولیان[3]

[1]دانشجوی کارشناسی ارشد گروه مهندسی کامپیوتر، دانشگاه بوعلی سینا، همدان

[2]استادیار گروه مهندسی کامپیوتر، دانشگاه بوعلی سینا، همدان

[3]استادیار گروه مهندسی کامپیوتر، دانشگاه بوعلی سینا، همدان

پست الکترونیک : mehrdad.hmt@gmail.com

چکیـــــده

مدل‌های موجود در روش‌های متداول برچسب‌گذاری تصاویر، نیازمند وجود دادگان از پیش تهیه شده بیشماری هستند. این دادگان یا از فضای مجازی تهیه می‌شوند که احتمال وجود نویز در آن‌ها بسیار بالاست و یا باید توسط افراد خبره تهیه شوند که بسیار زمان‌بر و پرهزینه است. به همین دلیل روش **"برچسب‌گذاری تصاویر بدون نمونه آزمایشی"** ارائه شده که در سال‌های اخیر توجه بسیاری از محققین را به خود جلب کرده است. ما در این پژوهش قصد داریم با استفاده از شبکه‌های عصبی بازگشتی و همچنین مدل زبانی word2vec، به دادگان بدون نمونه آموزشی، برچسب اختصاص دهیم.

واژه های کلیدی: برچسب‌گذاری، تصاویر بدون نمونه آزمایشی، شبکه عصبی بازگشتی

مقدمه

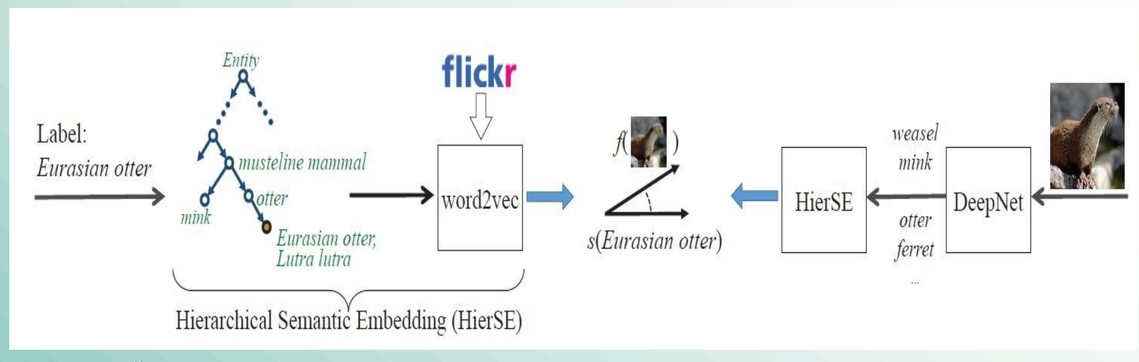
امروزه افراد برای یافتن تصاویر مورد نظرشان در فضاهای مجازی، از برچسب‌ها استفاده می‌کنند. روش‌های متداول برچسب‌زنی تصاویر، از شبکه‌های عصبی convolutional که بر روی میلیون‌ها داده از پیش برچسب‌گذاری شده آموزش دیده‌اند؛ استفاده می‌کنند که این روش‌ها در مواجهه با برچسب‌های نوین که هیچ نمونه آموزشی برای آن‌ها موجود نیست به مشکل برمی‌خورند. برای حل این مشکل روش برچسب‌زنی تصاویر بدون نمونه آموزشی ارائه شده است. طرح کلی این روش به این صورت است که یک **لایه میانی** به نحوی طراحی می‌شود که هم تصاویر و هم برچسب‌ها به این لایه نگاشت می‌شوند. به این ترتیب حتی برچسب‌های جدید نیز می‌توانند به این لایه نگاشت شوند. سپس با الگوریتم‌ها مختلف و بررسی ارتباط و شباهت‌ها موجود بین تصاویر و برچسب‌های نگاشت شده، می‌توان برچسب‌های جدید را به تصاویر نسبت داد.

برای نگاشت تصاویر و برچسب‌ها به لایه میانی مورد نظر، روش‌های متعددی وجود دارد. می‌توان برای نگاشت تصاویر از شبکه‌های از پیش یادگیری شده برای استخراج ویژگی تصاویر استفاده کرد. همچنین برای نگاشت برچسب‌ها نیز می‌توان از مدل‌های زبانی مختلف که توسط پایگاه‌داده‌های متنی همچون flickr، Wikipedia و ... یادگیری شده‌اند؛ استفاده کرد.

سابقه انجام پژوهش

در زمینه برچسب‌زنی تصاویر بدون نمونه آموزشی کارهای متنوعی انجام شده‌است. به طور کلی این تحقیقات را می‌توان به شکل زیر طبقه‌بندی نمود:

- نگاشت مستقیم تصاویر و برچسب‌ها به یک لایه میانی (DeViSE) [1]
- نگاشت غیرمستقیم تصاویر و نگاشت مستقیم برچسب‌ها به لایه میانی (ConSE) [2]
- برچسب‌گذاری تصاویر بدون نمونه آموزشی با استفاده از ساختار سلسله مراتبی WordNet [3]



شکل ۱: ساختار مدل مبتنی بر ساختار سلسله مراتبی WordNet برای برچسب زنی تصاویر بدون نمونه آموزشی [۳]

مجموعه داده

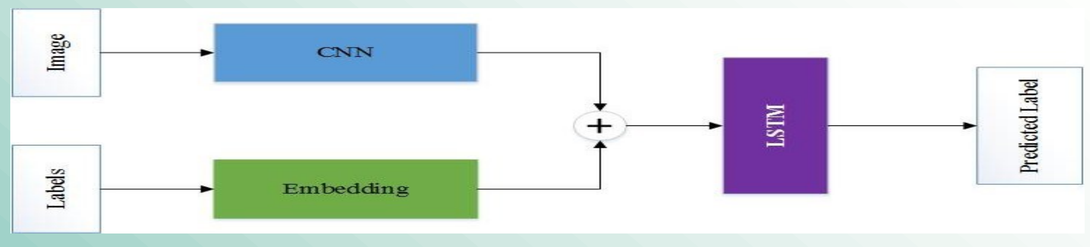
از میان پایگاه‌داده‌های موجود، آن‌هایی که به ازای هر تصویر چند برچسب دارند؛ برای برچسب‌زنی تصاویر بدون نمونه آموزشی مناسب‌تر می‌باشند.

جدول ۱: اطلاعات سه پایگاه‌داده Cifar100، ImageNet2010 و ImageNet2012

DataSet	Depth	Coarse NO.	Fine .No.
Cifar100	2	20	100
ImageNet 2010	2	143	387
ImageNet 2012	9	860	1000

راه حل پیشنهادی و کارهای آتی

ما در این پژوهش قصد داریم ابتدا با استفاده از شبکه‌های عصبی بازگشتی، به تصاویر پایگاه‌داده Cifar100، *برچسب‌های مرحله آموزش را اختصاص داده و سپس با استفاده از یک مدل زبانی یادگیری شده توسط flickr، برچسب‌های نوین اما مشابه با برچسب‌های تخمین زده شده را بیابیم. به این ترتیب هم تاثیر شبکه‌های بازگشتی بر روی برچسب‌گذاری تصاویر و هم نتیجه آموزش مدل زبانی توسط flickr در کار ما مشاهده خواهد شد.*



شکل ۲: ساختار مدل مبتنی بار CNN+RNN. در این مدل از شبکه عصبی بازگشتی LSTM برای برچسب زنی تصاویر استفاده شده است.

بحث و نتیجه‌گیری

در مقایسه با نتایج کارهای پیشین که با استفاده از مدل‌های گوناگون تنها قادر به تخمین یک برچسب برای برچسب‌گذاری تصاویر بودند؛ استفاده از شبکه LSTM برای برچسب‌گذاری تصاویر، امکان تخمین چندین برچسب به صورت همزمان را فراهم می‌سازد. همچنین با توجه به این که پایگاه‌داده flickr متشکل از برچسب‌های فراوانی است که کاربران به تصاویر اختصاص داده‌اند؛ نتیجه حاصل از مدل زبانی یادگیری شده توسط این پایگاه‌داده به مراتب بهتر از سایر پایگاه‌داده ها است.

منابع

[1] Andrea Frome et al. "DeViSE: A Deep Visual-Semantic Embedding Model", Advances in Neural Information Processing Systems 26 (NIPS 2013)

[2] Mohammad Norouzi et al, "Zero-Shot Learning by Convex Combination of Semantic Embeddings", arXiv preprint arXiv:1312.5650, 2013.

[3] X. Li, S. Liao, W. Lan, X. Du, and G. Yang. "Zero-shot image tagging by hierarchical semantic embedding". In Proc. of SIGIR , 2015.

تقدیر و تشکر:

با سپاس فراوان از زحمات استاد گرانقدر جناب آقای دکتر محرم منصوری‌زاده