یادگیری فعال و کاربرد آن در برچسب‌زنی دستوری کلمات

محمد امین مهرعلیان[۱] ، شهرام خدیوی[۲]
[۱]دانشکده‌ی مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر
mehralian@aut.ac.ir

[۲]دانشکده‌ی مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر
khadivi@aut.ac.ir

چکیده

یادگیری ماشین با رویکرد با ناظر امروزه پایه‌ی بسیاری از فعالیت‌های مختلف در حوزه پردازش زبان طبیعی است. اگرچه این روش‌ها به موفقیت‌هایی دست یافته‌اند اما در مقابل نیازمند فراهم شدن حجم زیادی از داده آموزشی توسط یک تفسیر کننده مانند انسان است، که گاها هزینه‌های بالایی را در بر خواهد داشت. علاوه بر این در بیشتر روش‌های یادگیری با ناظر ترتیب انتخاب نمونه‌های آموزشی بر اساس تصادف صورت می‌گیرد، در مقابل برای برطرف کردن مشکلات مذکور یادگیری فعال مطرح می‌شود که در آن به شکلی تکرار شونده نمونه‌هایی با بیشترین اثر مطلوب بر فرآیند آموزش انتخاب می‌شوند.

نتایج آزمایشات نشان می‌دهد در آموزش یک مدل برچسب‌زنی دنباله فارسی بر اساس پیکره متنی زبان فارسی تنها با استفاده از ۹,۳۶٪ از کل داده‌های آموزشی به دقت برچسب زنی تا ۹۶,۲۸٪ رسید و این در حالی است که دقت برچسب زنی با بکارگیری کل نمونه‌ها ۹۶,۴۵٪ می‌باشد.

کلمات کلیدی

یادگیری فعال، برچسب‌زنی دستوری کلمات، پیکره متنی زبان فارسی

۱- مقدمه

استفاده از روش‌های معمول یادگیری باناظر در کاربردهایی چون پردازش زبان طبیعی نیازمند حجم بالایی از داده‌های آموزش است. در این رویکرد نمونه‌های آموزشی به صورت تصادفی انتخاب می‌شود و از آنجا که یادگیری بر اساس نمونه‌هایی که قبلا مشابه آن‌ها به سیستم ارائه شده است، اثر آموزشی کمتری برای مدل خواهد داشت، بنابراین بخش اعظمی از داده‌های آموزشی اثرگذاری کمی بر فرآیند یادگیری دارند، و این درحالی است که برچسب زدن هریک از این نمونه‌ها توسط انسان **هزینه**‌هایی در بردارد. علاوه بر این، اگر آموزش بر اساس نوع خاصی از نمونه‌ها زیاد تکرار شود مشکل overfiting رخ خواهد داد، بنابراین با محدود کردن نمونه‌هایی که مدل قبلا به اندازه کافی بر اساس آن‌ها آموزش دیده است مشکل overfiting نیز به صورت خودکار حل خواهد شد.

در مقابل یادگیری فعال با بکارگیری یک رویکرد هدفمند تلاش می‌کند با کمترین نمونه‌ی آموزشی و متعاقبا کمترین هزینه برچسب‌زنی بیشترین اثر یادگیری را اعمال کند. در این روش تنها نمونه‌هایی برای برچسب زدن توسط انسان و افزودن آن به مجموعه‌ی آموزشی انتخاب می‌شوند که سیستم در مورد آنها اطلاعات کافی را ندارد و به بیانی دیگر، قبلا بر اساس نمونه‌های مشابه آن آموزش ندیده باشد. استفاده از این راهکار در حوزه‌هایی چون پردازش زبان طبیعی که به حجم بالایی از داده نیاز داشته و برچسب زدن آن‌ها پرهزینه است، ضروری به نظر می‌رسد.

کاربرد این رویکرد در بسیاری از حوزه‌های مربوط به برچسب‌گذاری دنباله (sequence labeling) توسعه فراوانی یافته است. به عنوان نمونه در [1] از یادگیری فعال برای برچسب زدنی دستوری کلمات(POS)، استفاده شده است و نویسنده با آزمایش روش‌های

مختلف نمونه‌برداری در یادگیری فعال به مقایسه آن‌ها پرداخته است، علاوه براین وی روشی جدید برای نمونه‌برداری در این حوزه را معرفی کرده است که مبتنی بر فرکانس تکرار کلمات در پیکره عمل می‌کند. در [2] برای تعیین گروه اسمی از یادگیری فعال با نمونه‌برداری بر اساس آنتروپی بهبود یافته استفاده شده است که در آن از معیار تنوع در پیکره استفاده شده است. در [3] برای تعیین گروه اسمی بجای استفاده از یادگیری فعال در سطح جمله از یادگیری فعال در سطح کلمه استفاده شده است که هزینه برچسب‌زنی را تا حدی کاهش داده است. در [4] از یادگیری فعال برای استخراج اطلاعات (information extraction) استفاده شده است، نویسنده در این مقاله از مدل مخفی مارکوف مبتنی بر یادگیری فعال و از نمونه‌برداری بر اساس حداکثر حاشیه استفاده کرده است.

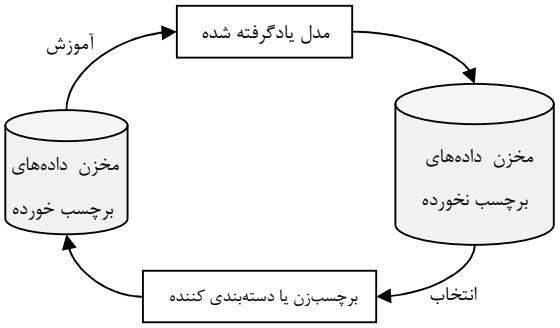
در این مقاله این رویکرد را برای برچسب‌زنی دستوری کلمات در فارسی بکارگرفته شده است و نتایج نهایی نشان می‌دهد از مجموع کل داده‌های آموزش تهیه شده برای این منظور در [9] تنها با استفاده از یک دهم داده‌ها و متعقبا کاهش هزینه‌ی به یک دهم برای تولید داده‌ی آموزشی می‌توان به دقت برچسب زنی مشابهی رسید.

۲- یادگیری فعال

یادگیری فعال یک نوع یادگیری با نظارت است که یادگیرنده در آن با مجموعه‌ای از داده‌های کنترل شده آموزش می‌بینند. بدین معنا که یک سیستم کنترلی با استفاده از مدلی که با اطلاعات اولیه تخمین زده شده است، مشخص می‌کند کدام داده توسط یک یاد دهنده مثل انسان که دارای اطلاعات در زمینه یادگیری سیستم است، برچسب زده شود. در هرگام سیستم با استفاده از داده‌های جدید کامل‌تر می‌شود و میزان دقت خود را افزایش می‌دهد تاجایی که در مراحل پایانی اجرای آن تقریبا در مورد دسته‌بندی داده‌ها به قطعیت کامل می‌رسد.

در کل، هدف پیدا کردن یک دسته‌بندی کننده تا حد امکان دقیق، به نحوی است که حداقل مقدار داده ممکن از طرف یاددهنده برچسب زده شود. تفاوت اصلی در این سیستم یادگیری با روش‌های معمول با نظارت دیگر در این است که در مفهوم یادگیری منفعل که در برابر یادگیری فعال تعریف می‌شود، آموزش بر اساس یک نمونه‌برداری تصادفی صورت می‌گیرد.

فرآیند اجرای یادگیری فعال در شکل (۱) نشان داده شده‌است. همانطور که دیده می‌شود داده‌های موجود به دوبخش داده‌های برچسب خورده و داده‌های برچسب نخورده تقسیم می‌شود و در هر مرحله آموزش، بخشی از داده‌های برچسب نخورده انتخاب شده و برای برچسب زدن به سیستمی که وظیفه برچسب‌زنی را دارد (مثل انسان) ارسال می‌شود و مجددا مدل بر اساس داده‌های جدید آموزش می‌بینند. بنابراین مهمترین بخش در این نوع یادگیری فرآیند انتخاب نمونه یا ساخت پرس‌وجو (Query) به منظور مشخص کردن مجموعه‌ای از داده‌های با ارزش برای آموزش در هر مرحله است. برای این منظور روش‌های متفاوتی بکارگرفته شده است که هرکدام بسته به شرایط و نوع مدل ویژگی‌هایی را دارند که در بخش بعد در مورد آن‌ها توضیح خواهیم داد.



شکل (۱) : فرآیند اجرای یادگیری فعال

۲-۱- نمونه‌برداری بر اساس عدم قطعیت

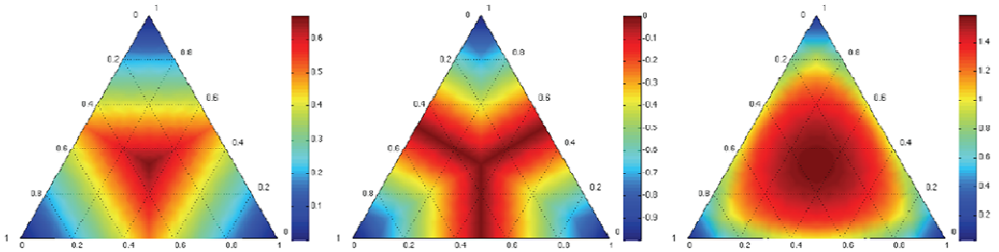
یکی از معمول‌ترین روش‌های پرس‌وجو استفاده از نمونه‌برداری بر اساس عدم قطعیت است، در این روش پرس‌وجو نمونه‌هایی که کمترین میزان قطعیت برای برچسب متناظر با آن‌ها وجود دارد، انتخاب می‌شوند. به عنوان نمونه در مدل احتمالاتی برای یک مساله دسته‌بندی دودویی، این پرس‌وجو نمونه‌ای را انتخاب خواهد کرد که احتمال ثانویه (posterior) برای آن فاصله کمی با ۰,۵ داشته باشد.

در حالت کلی برای یک مساله چند کلاسه معمول‌ترین معیاری که برای اندازی‌گیری این عدم قطعیت استفاده می‌شود، بکارگیری معیار *کمترین اطمینان* است.

$$x^*_{LC} = \text{argmax}_x \; 1 - p_\theta(\hat{y}|x) \qquad (۱)$$

که در آن $\hat{y} = \text{argmax}_y \, p_\theta(y|x)$ بیشترین احتمال ثانویه، برای کلاس برچسب متناظر با ورودی x بر اساس مدل θ است. معنای دیگر معیار عدم قطعیت، انتخاب نمونه‌هایی است که برچسب آنها از نظر مدل معلوم نیست.

اگرچه معیار کمترین اطمینان به عنوان یکی از معمول‌ترین معیاری‌ها برای اندازه‌گیری عدم قطعیت بکار می‌رود اما به هرحال این استراتژی تنها اطلاعات مربوط به محتمل‌ترین برچسب را در نظر می‌گیرد و اطلاعات مربوط به توزیع سایر برچسب‌ها نادیده گرفته می‌شود. برای برطرف کردن این نقیصه معیار دیگری بنام *نمونه‌برداری بر اساس*



شکل (۲): نمودار حرارتی برای سه معیار مختلف جهت اندازه‌گیری میزان عدم قطعیت در نمونه‌ها (الف) عدم قطعیت با استفاده از معیار کمترین اطمینان (ب) عدم قطعیت با استفاده از معیار مبتنی بر حاشیه (پ) عدم قطعیت با استفاده از معیار آنتروپی [7]

حاشیه معرفی می‌شود که بر اساس آن برای یک مساله چند کلاسه رابطه زیر را خواهیم داشت.

$$x_M^* = \text{argmin}_x\, p_\theta(\hat{y}_1|x) - p_\theta(\hat{y}_2|x) \qquad (۲)$$

که در آن $\hat{y}_1$، $\hat{y}_2$ اولین و دومین محتمل‌ترین کلاس برچسب برای ورودی x بر اساس مدل θ است. این معیار با مشارکت دادن دومین محتمل‌ترین کلاس مشکل معیار قبلی را تا حدی برطرف می‌کند، بدین صورت که نمونه‌هایی که دارای حاشیه زیادی هستند (فاصله‌ی دو کلاس محتمل‌تر دور هستند) به راحتی دسته‌بندی می‌شوند و طبیعتا در پرس‌وجو وارد نخواهند شد ولی نمونه‌هایی با حاشیه کم، از آنجا که احتمال تعلق آنها به دو کلاس متفاوت با احتمال بالایی وجود دارد، بنابراین در مورد آنها با ابهام روبرو هستیم و به عنوان نمونه‌هایی با عدم قطعیت بالا شناخته می‌شوند.

به هرحال این معیار نیز جز دو کلاس محتمل، توزیع سایر کلاس‌ها را در نظر گرفته نمی‌گیرد. در این صورت از یک معیار عمومی‌تر بنام *آنتروپی* استفاده می‌کنیم که بر اساس آن داریم:

$$x_H^* = \text{argmax}_x - \sum_i p_\theta(y_i|x) \log p_\theta(y_i|x) \qquad (۳)$$

که در آن y_i در بازه تمام حالات ممکن از برچسب‌ها قرار دارد.

برای یک مساله دسته‌بندی دودویی همه روش‌های فوق معادل با یکدیگر خواهند بود و نمونه‌ای انتخاب خواهد شد که احتمالات ثانویه‌ی آن فاصله کمی را با ۰٫۵ داشته باشد. اما به هرحال معیار آنتروپی گزینه مناسب‌تری برای ساختارهای پیچیده است.

شکل (۲) نمودار حرارتی مربوط به این سه روش را برای یک مساله ۳ کلاسه نشان می‌دهد که هر گوشه این مثلث نماینده یکی از این کلاس‌هاست. اضلاع بین هر دو کلاس نشان دهنده احتمال تعلق به کلاس، از یک سر ضلع، به کلاس موجود، در سر دیگر آن است. در این نمودار هر نقطه متناسب با میزان تیرگی آن از میزان عدم قطعیت بیشتری برخوردار است. نتایج تجربی حاصل از مقایسه این سه روش نشان می‌دهد که بهترین استراتژی بسته به نوع کاربرد آن متفاوت خواهد بود

۳- یادگیری فعال در برچسب‌زنی دنباله‌ها

استفاده از یادگیری فعال بیشتر در زمینه دسته‌بندی سابقه دارد و بکارگیری آن در حوزه‌ی برچسب‌زنی دنباله، کمتر مورد توجه بوده است. بیشتر کارهای انجام شده در این زمینه نیز مبتنی بر نمونه‌برداری بر اساس عدم قطعیت است [5]. به عنوان نمونه در [6] نویسنده از رویکرد کمترین اطمینان استفاده کرده است و در [4] از روش حداکثر حاشیه، برای استخراج اطلاعات (information extraction) استفاده شده است، همچنین نویسنده در [2] با بکارگیری روش حداقل آنتروپی در ۴ شکل متفاوت از آن به نتایج مناسبی رسیده است.

در این مقاله نیز از رویکرد نمونه‌برداری بر اساس عدم قطعیت استفاده شده است و نتایج روش‌های مختلف مطرح در این زمینه مورد بررسی قرار گرفته است.

۴- آزمایشات

در این آزمایشات از الگوریتم Maximum Entropy برای ایجاد مدل برچسب‌زن استفاده شده است. داده‌های آموزشی بکارگرفته شده از پیکره متنی زبان فارسی [9] است که ۸۰٪ آن برای آموزش، و ۲۰٪ باقی مانده برای اندازه گیری دقت مدل استفاده شده است.

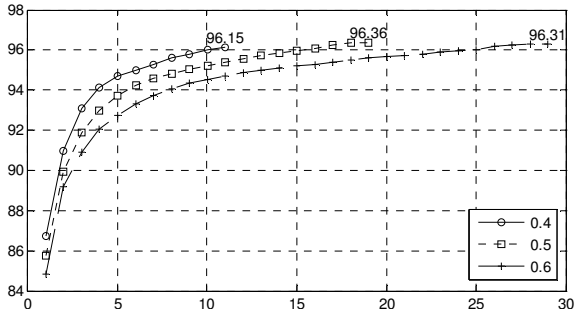
لازم به ذکر است که در این آزمایشات هدف صرفا بررسی اثر یادگیری فعال بر فرآیند آموزش در این مساله خاص بوده است و نه بررسی عملکرد یک سیستم مبتنی بر Maximum Entropy برای برچسب زنی دستوری کلمات در فارسی.

در ابتدای فرآیند آموزش ابتدا مدل با استفاده از مجموعه‌ی اولیه‌ای شامل ۲۵ جمله آموزش دیده است، سپس نمونه‌های جدید برای برچسب‌گذاری به مدل ارائه می‌شوند، اگر نمونه بر اساس معیار مورد نظر برای آزمایش و حد آستانه تعیین شده برای آن معیار با قطعیت مناسبی برچسب گذاری نگردد، به عنوان یک نمونه با ارزش برای یادگیری کنار گذاشته می‌شود و نمونه بعدی مورد بررسی قرار می‌گیرد. این فرآیند تا رسیدن تعداد نمونه‌های با ارزش، به مقدار مشخصی ادامه

می‌یابد و سپس مدل مجددا با نمونه‌های جدید به روز می‌شود. گام‌های بعدی نیز همین روند را تا اتمام تمام نمونه‌های یادگیری طی می‌کنند.

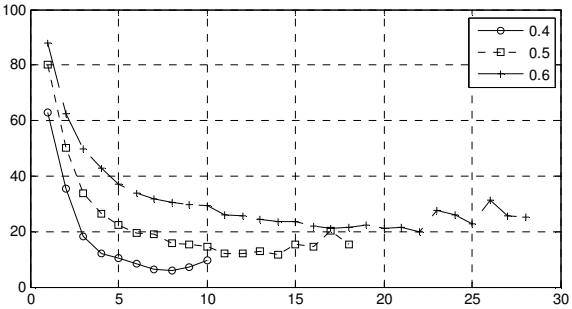
۴-۱- یادگیری فعال براساس کمترین اطمینان

در این آزمایش یادگیری فعال بر اساس پرس‌وجوی مبتنی بر کمترین اطمینان با مقادیر مختلف از حد آستانه مورد بررسی قرار گرفته است. شکل (۳) دقت برچسب زنی سیستم را در هر گام مشخص می‌کند. همانطور که مشاهده می‌شود سیستم در گام‌های ابتدایی به سرعت به بلوغ رسیده و در گام‌های انتهایی همگرا شده است. و در استفاده از مقادیر بزرگتر برای این پارامترها همگرایی با سرعت کمتری حاصل می‌شود.



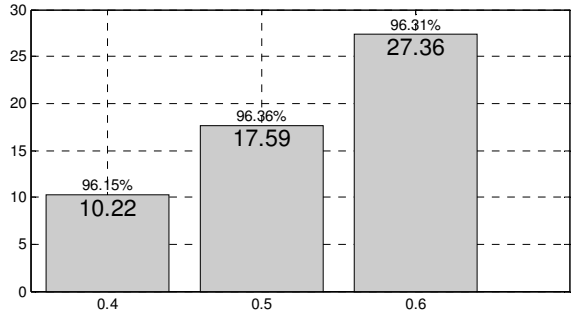
شکل (۳) : دقت برچسب زنی در هرگام برای یادگیری فعال براساس کمترین اطمینان

در شکل (۴) درصد داده‌های با قطعیت پایین برای مدل، در هر گام نشان داده شده است (درصد نمونه‌های با ارزش برای یادگیری). همانگونه که مشاهده می‌شود در گام‌های پایانی که دقت سیستم به همگرایی می‌رسد این نمودار نیز شروع به نوسان می‌کند. در کل بررسی نتایج در گام‌های ابتدایی آموزش نشان دهنده اثر یادگیری فعال بر روند آموزش است، اما ثابت شدن میزان اثربخشی در گام‌های پایانی نشان می‌دهد مدل از تمام ظرفیت یادگیری خود استفاده کرده است و با افزایش تعداد گام‌ها بازهم چندان قادر به افزایش دقت خود نیست.



شکل (۴) : درصد نمونه‌های با ارزش در هر گام، برای یادگیری فعال براساس کمترین اطمینان

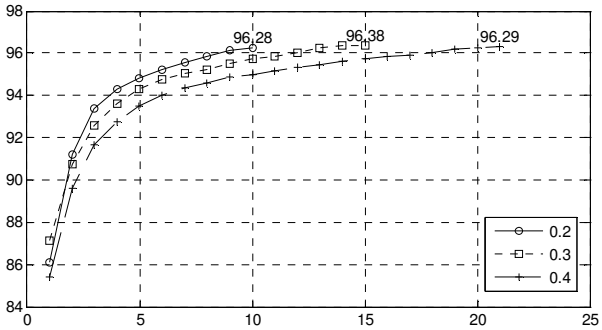
و در نهایت در شکل زیر درصد داده‌های آموزشی مورد استفاده برای آموزش مدل بر اساس مقادیر مختلف پارامتر مشاهده می‌شود که متناسب با تعداد گام‌های یادگیری‌فعال افزایش می‌یابد. همانطور که می‌بینید در این روش با درنظر گرفتن حدآستانه ۰,۴ برای محتمل‌ترین کلاس تنها نیاز به آموزش بر اساس ۱۰,۲۲٪ از داده‌های آموزشی است. در این نمودار عدد بالای ستون‌ها بیانگر دقت برچسب زنی و عدد پایینی نشان دهنده درصد داده‌های استفاده شده برای آموزش است.



شکل (۵): درصد داده‌های استفاده شده برای آموزش مدل به ازای پارامترهای مختلف در یادگیری فعال براساس کمترین اطمینان

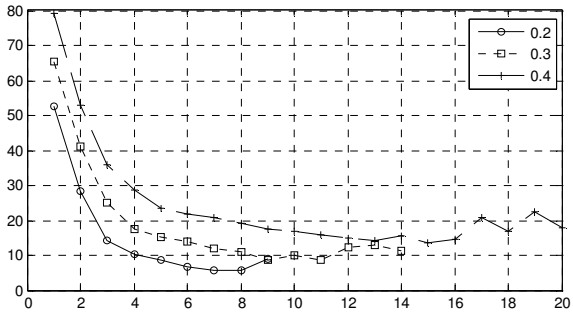
۴-۲- یادگیری فعال براساس حداکثر حاشیه

در این آزمایش یادگیری فعال براساس حداکثر حاشیه با مقادیر ۰,۴، ۰,۳ و ۰,۲ برای پارامتر آن آزمایش شده است. از آنجا که تحلیل‌های ارائه شده در آزمایش قبلی عینا در این آزمایش نیز صدق می‌کند بنابراین تنها به ارائه نتایج عددی و نمودارها اکتفا می‌کنیم.



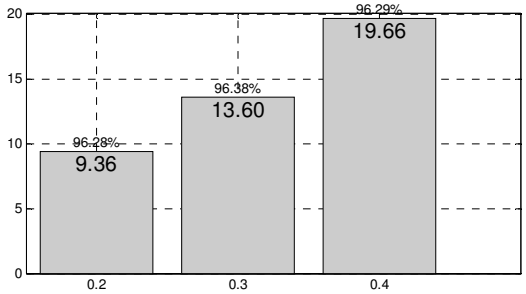
شکل (۶) : دقت برچسب زنی در هرگام، برای یادگیری فعال براساس حداکثر حاشیه

مقایسه نمودار شکل (۶) و نمودار بدست آمده در آزمایش قبل نشان می‌دهد در این روش همگرایی با سرعت بیشتری ایجاد شده است. به طوری که در ۱۰ گام به دقت ۹۶,۲۸٪ رسیده‌ایم در حالی که در روش قبل در ۱۱ گام اولیه دقت ۹۶,۱۵٪ بوده است.



شکل (۷) : درصد نمونه‌های با ارزش در هر گام، برای یادگیری فعال براساس حداکثر حاشیه

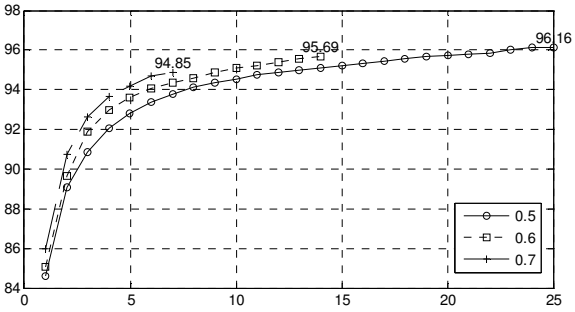
نمودار زیر نشان می‌دهد در این روش علاوه بر افزایش میزان دقت برچسب زنی ، درصد کمتری از داده‌ها نیز مورد استفاده بوده است.



شکل (۸) : درصد داده‌های استفاده شده برای آموزش مدل به ازای پارامترهای مختلف در یادگیری فعال براساس حداکثر حاشیه

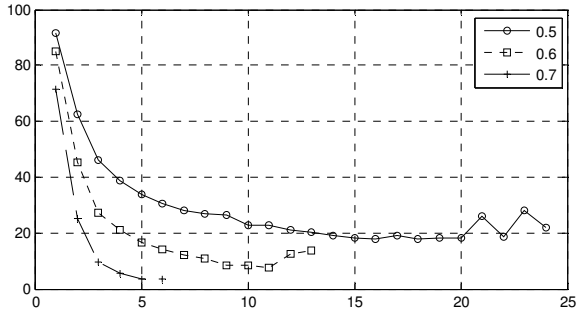
۱-۱ یادگیری فعال براساس حداقل آنتروپی

در این آزمایش یادگیری فعال براساس حداقل آنتروپی با مقادیر ۰,۵، ۰,۶ و ۰,۷ برای پارامتر آزمایش شده است.



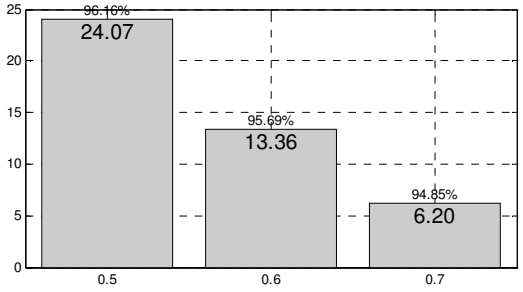
شکل (۹) : دقت برچسب زنی در هرگام، برای یادگیری فعال براساس حداقل آنتروپی

در این روش علاوه بر همگرایی دقت برچسب زنی نیز کاهش یافته است، به عنوان نمونه در این روش برای رسیدن به دقت ٪۹۶,۱۶ ، ۲۵ گام طی شده است اما برای رسیدن به همین میزان دقت در روش قبل تنها ۹ گام نیاز بوده است.



شکل (۱۰): درصد نمونه‌های با ارزش در هر گام برای یادگیری فعال براساس حداقل آنتروپی

حجم داده مورد نیاز برای آموزش نیز وضعت مشابهی دارد، در این روش برای پارامتر ۰,۶ با استفاده از ٪۱۳,۳۶ از داده‌ها به دقت ٪۹۵,۶۹ رسیده‌ایم در حالی که در روش قبل با پارامتر ۰,۳ با استفاده از ٪۱۳,۶ داده‌ها به دقت ٪۹۶,۳۸ رسیده‌ایم.



شکل (۱۱) : درصد داده‌های استفاده شده برای آموزش مدل به ازای پارامترهای مختلف در یادگیری فعال براساس حداقل آنتروپی

۵- نتیجه گیری

با نگاهی به مجموع بهترین نتایج حاصل از هر روش در جدول زیر می‌توان دریافت، استفاده از یادگیری فعال در این مساله، زمانی بیشترین کارایی را از خود نشان می‌دهد که از معیار حداکثر حاشیه استفاده شده باشد.

جدول (۱) : بهترین نتایج حاصل از هر روش

روش	دقت مدل	درصد داده استفاده شده
کمترین اطمینان	٪۹۶,۱۵	٪۱۰,۲۲
حداکثر حاشیه	٪۹۶,۲۸	٪۹,۳۶
حداقل آنتروپی	٪۹۶,۱۶	٪۲۴,۰۷
دقت برچسب زنی با استفاده از تمام نمونه‌های آموزشی ۹۶,۴۵		

بر این اساس برای آموزش یک مدل با دقت مناسب کافیست تنها نمونه‌هایی برای آموزش برچسب گذاری شوند که دو کلاس محتمل‌تر از سوی مدل با قطعیتی نزدیک بهم انتخاب شوند و به بیانی دیگر محتمل‌ترین برچسب متناظر با ورودی با فاصله کمی از دومین محتمل‌ترین برچسب، از سوی مدل انتخاب شده باشد.

مقایسه نتایج این روش با سایر روش‌ها با توجه به مرجع [8] نشان می‌دهد علی‌رغم استفاده از تنها یک‌دهم داده‌ها نتایج نهایی چندان دچار کاهش نشده‌اند.

جدول (۲) : مقایسه دقت برچسب زنی در این روش و سایر روش‌ها

روش مورد استفاده	درصد بازشناسی
TnT(Markov Model)	٪۹۶,۶۴
MBT(Memory Based)	٪۹۶,۴۲
MLE(Maximum Likelihood)	٪۹۶,۶۰
روش ما با ۹,۳۶٪ از داده‌ها	٪۹۶,۲۸
روش ما با استفاده از کل داده‌ها	٪۹۶,۴۵

مراجع

[1] E. Ringger, et al., "Active learning for part-of-speech tagging: Accelerating corpus annotation," in *Proceedings of the Linguistic Annotation Workshop*, 2007, pp. 101--108.

[2] S. Kim, Y. Song, K. Kim, J. W. Cha, and G. G. Lee, "MMR-based active machine learning for bio named entity recognition," in *Proceedings of the Human Language Technology Conference of the NAACL*, 2006, pp. 69--72.

[3] K. Tomanek and U. Hahn, "Semi-supervised active learning for sequence labeling," in , 2009, pp. 1039--1047.

[4] T. Scheffer, C. Decomain, and S. Wrobel, "Active hidden Markov models for information extraction," *Advances in Intelligent Data Analysis*, pp. 309--318, 2001.

[5] B. Settles and M. Craven, "An analysis of active learning strategies for sequence labeling tasks," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2008, pp. 1070--1079.

[6] A. Culotta and A. McCallum, "Reducing labeling effort for structured prediction tasks," in *PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE*, 2005, pp. 746-751.

[7] B. Settles, "Active Learning Literature Survey," techreport, 2009,.

[8] F. Raja, et al., "Evaluation of part of speech tagging on Persian text," in , 2007.

[9] M.Bijankhan, J. Sheikhzadegan, M. Bahrani, M. Ghayoomi, "lessons from building a persian written corpus:peykare", language resources & evaluation, 2010.