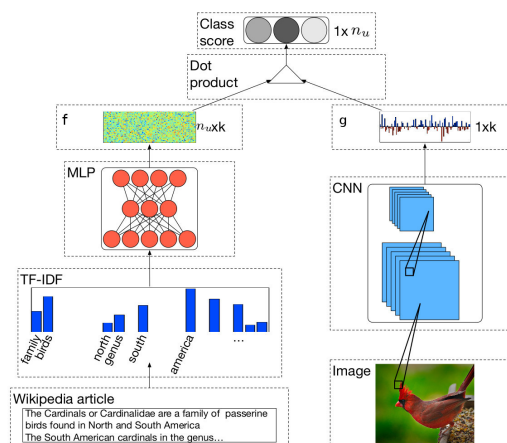


است. محدودیت $\zeta_i - (w^T x_i) \geq 0$ بر اساس فرض عدم تعلق نمونه‌های آزمون به کلاس‌های دیده‌شده تعریف شده است و اجبار می‌کند که تمامی نمونه‌های دیده‌شده باید در طرف منفی دسته‌بند خطی w قرار گیرند. نویسندگان این پژوهش، روش خود را با استفاده از تکنیک هسته ^{۲۹} برای دسته‌بندهای غیرخطی نیز توسعه داده‌اند [۳۵].

۸-۲ نداشت به یک فضای میانی

در برخی روش‌ها هر دوی نگاشت‌های ϕ و θ ، معرفی شده در ابتدای فصل با توجه به داده‌ها یاد گرفته می‌شوند و در نتیجه فضای مشترک مورد استفاده نه فضای تصاویر و نه فضای توصیف‌هاست؛ بلکه فضای ثالثی است. این فضای میانی در برخی از روش‌ها یک فضای با بعد کمتر است و تعبیر معنایی برای آن موجود نیست. در برخی روش‌های دیگر، فضای میانی را با بعد n_s یعنی تعداد دسته‌های دیده شده در نظر گرفته‌اند و تعبیر معنایی برای آن ارائه شده است. این فضای میانی بر اساس توصیف دسته‌ها و نمونه‌های دیده نشده بر اساس شباهت آن‌ها با دسته‌های دیده شده استوار است.



شکل ۲-۲: شبکه مورد استفاده برای یادگیری توأم نگاشت تصاویر و توصیف‌ها که یک شبکه عصبی ژرف با دو ورودی است. ورودی اول از نوع تصویر است و ابتدا با یک شبکه پیچشی سپس با چند لایه چگال به فضایی k -بعدی می‌رود. ورودی دوم که یک مقاله از ویکی‌پدیای انگلیسی است پس از تبدیل به نمایش برداری به صورت $tf-idf$ با چند لایه با اتصالات چگال پردازش شده و به فضایی k -بعدی می‌رود. در نهایت امتیاز تعلق تصویر به دسته‌ی متن با ضرب داخلی این دو نگاشت تعیین می‌شود [۱].

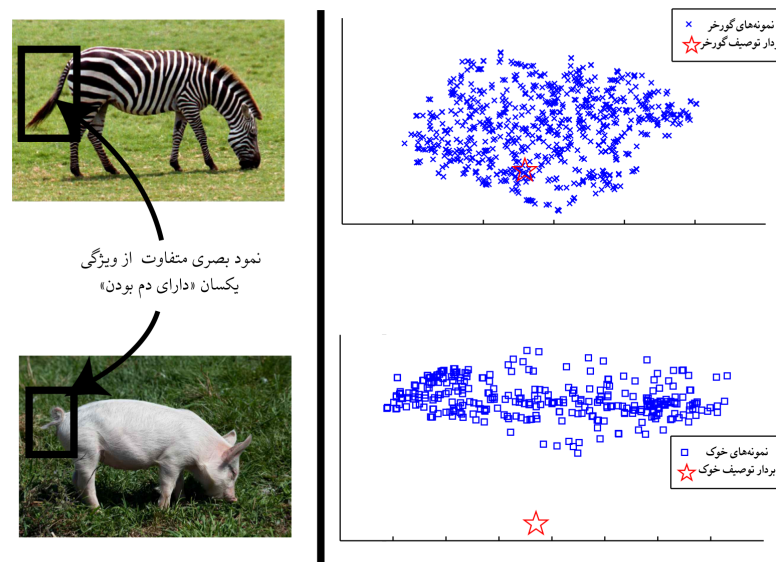
در [۱] از شبکه‌های عصبی ژرف برای یادگیری توأم نگاشت‌های ϕ و θ استفاده شده است. نمای کلی شبکه مورد استفاده در این روش در تصویر ۲-۲ نشان داده شده است. توصیف‌های متنی و ویژگی‌های بصری دو ورودی جداگانه

^{۲۹}kernel trick

۹-۲ روش‌های نیمه‌نظارتی

در این بخش به بررسی روش‌های نیمه‌نظارتی می‌پردازیم. این روش‌ها از نظر نوع نگاشت‌های مورد استفاده در یکی از دسته‌های قبلی قابل بیان بودند ولی با توجه به این که روش پیشنهادی ما نیز نیمه‌نظارتی است، برای پررنگ‌تر شدن نحوه‌های استفاده از داده‌های آزمون در جریان آموزش این دسته را به طور جداگانه مورد بررسی قرار می‌دهیم.

در [۳۸] برای نخستین بار مشکل جابجایی دامنه^{۳۹} معرفی شد. این مشکل که در شکل ۲-۳ قابل مشاهده است به متفاوت بودن خواص صفت‌ها برای دسته‌های مختلف اشاره می‌کند. برای مثال صفت راه‌راه بودن برای دو حیوان گورخر و ببر از نظر بصری خواص متفاوتی دارد و یادگیری یک دسته‌بند برای تشخیص راه‌راه بودن با استفاده از تصاویر گورخر در تشخیص وجود و یا عدم وجود این صفت در تصویر ببر ضعیف خواهد بود.

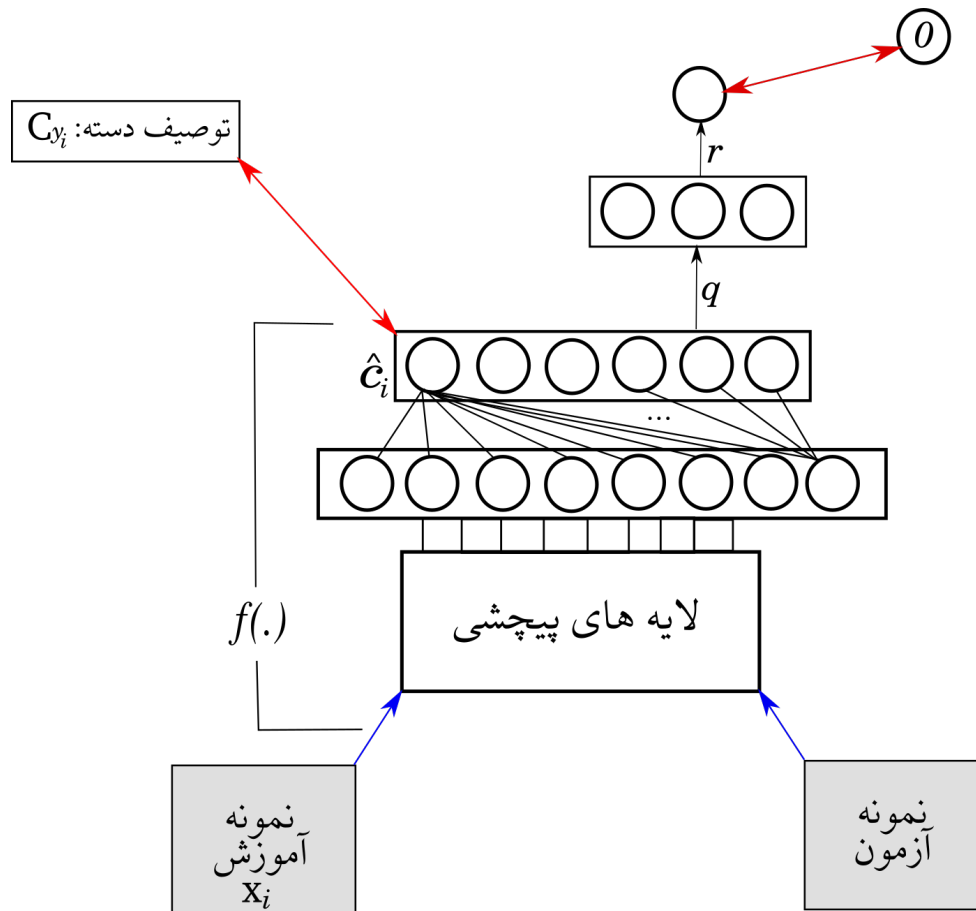


شکل ۲-۳: مشکل جابجایی دامنه بین دو دسته‌ی دیده شده (گورخر) و دیده نشده (خوک) نمایش داده شده است. صفت یکسان «دارای دم بودن» در این دو دسته دارای دو نمود بصری متفاوت است (سمت چپ) و نگاشت یادگرفته شده برای بردن این صفت به فضای مشترک برای دسته‌ی دیده نشده عمل‌کرد ضعیف‌تری نسبت به دسته‌ی دیده شده به نمایش می‌گذارد (سمت راست) [۳۸].

در [۳۸] برای حل این مشکل دو تکنیک به کار گرفته شده است. ابتدا یافتن نمایش مشترک برای سه دامنه‌ی تصاویر، بردار صفت و بردار نام دسته‌ها به صورت توأمان با استفاده از CCA^{۴۰} [۳۹] و سپس برچسب‌گذاری داده‌های بدون

^{۳۹}Domain shift problem

^{۴۰}Canonical Correlation Analysis

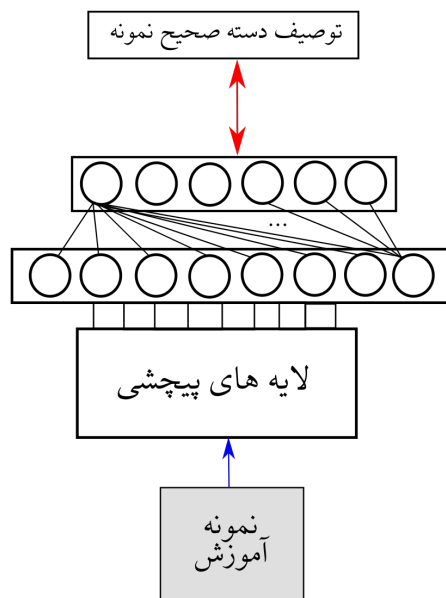


شکل ۳-۲: ساختار شبکه چند وظیفه‌ای پیشنهادی. فلش‌های آبی رنگ ورودی‌های شبکه را نشان می‌دهند و فلش‌های قرمز رنگ مقایسه خروجی شبکه با خروجی مورد انتظار را. خطوط سیاه رنگ اتصالات شبکه را نشان می‌دهند. زیر شبکه‌ی برگرفته شده از شبکه vgg و یک لایه‌ی با اتصالات چگال اضافه شده بین دو ورودی مشترک هستند. لایه‌های r و q مخصوص نمونه‌های آزمون هستند. خروجی لایه‌ی r همواره با مقدار صفر مقایسه می‌شود.

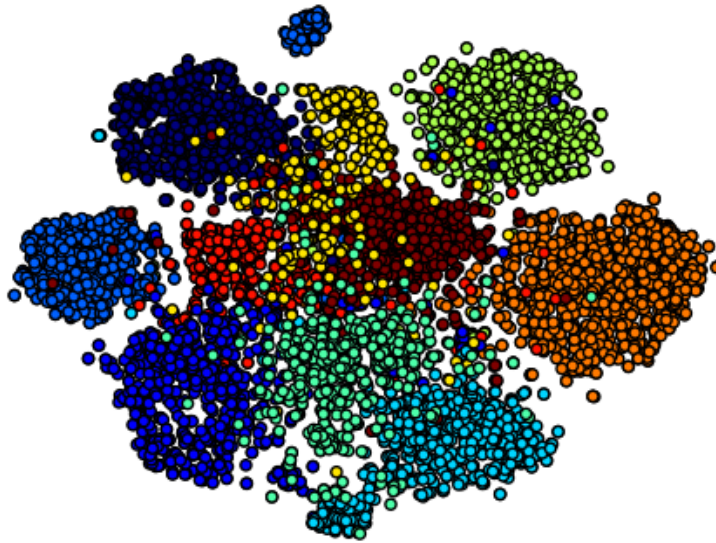
دیگری استخراج شده است. یعنی ابتدا یک روش دسته‌بندی احتمالی (در آن پژوهش روش IAP [۱۳] برای این کار انتخاب شده) به صورت مستقل روی مجموعه دادگان اجرا شده و احتمال‌هایی که برای انتساب هر نمونه به دسته‌های آزمون از آن روش بدست می‌آید بعنوان وزن‌های شباهت در نظر گرفته می‌شود. فاصله بردار صفت پیش‌بینی شده برای هر نمونه با توصیف دسته‌های آزمون متناسب با این وزن‌های شباهت جریمه می‌شود. ما در این بخش یک روش مبتنی بر شبکه‌های عصبی ژرف معرفی می‌کنیم که در آن نگاشتی غیرخطی و چندلایه از تصاویر به بردارهای صفت یادگرفته می‌شود. معیار یادگیری این نگاشت، پیش‌بینی صحیح صفت برای نمونه‌های آموزش (که بردار صفت صحیح برای آن‌ها مشخص است) و هم‌چنین نزدیک بودن حاصل نگاشت هر نمونه‌ی آزمون به توصیف یکی از دسته‌های دیده نشده است.

۳-۲-۳ یک مدل پایه برای مقایسه

برای روشن شدن تاثیر استفاده از اطلاعات بدون نظارت نمونه‌های آزمون در یادگیری بهتر نگاشت، قصد داریم در فصل آتی مدل ارائه شده را با یک مدل ساده برای پیش‌بینی صفت مقایسه کنیم که در این جا این مدل پایه را معرفی می‌کنیم. در این مدل ساده تنها از لایه‌های با اتصالات کامل بعد از استخراج ویژگی با لایه‌های پیچشی، برای پیش‌بینی صفت استفاده شده است. ساختار این مدل در تصویر ۳-۳ نمایش داده شده است. در این شبکه از یک یا چند لایه با اتصالات کامل بعد از لایه‌های پیچشی استفاده می‌شود. مشابه حالت قبل تابع فعال‌سازی برای مجموعه دادگانی که مقادیر توصیف دسته‌هایشان دودویی است تابع سیگموید، و برای مجموعه دادگانی که مقادیر بردارهای توصیف در آن‌ها مقادیر دلخواه حقیقی است تابع ReLU در نظر گرفته شده است. ابعاد لایه‌های با اتصالات کامل پایانی الزاماً برابر تعداد ابعاد بردارهای توصیف است و برای سایر لایه‌های با اتصالات کامل نیز همین تعداد ابعاد انتخاب شده است. مقایسه نتایج دقت دسته‌بندی بین مدل قبلی و این مدل در بخش ۴-۴ نشان‌دهنده تاثیر مثبت استفاده از اطلاعات بدون نظارت موجود در نمونه‌های آزمون است که باعث بهبود حداقل ۱۰ درصدی دقت دسته‌بندی می‌شود.



شکل ۳-۳: ساختار شبکه پایه. فلش آبی رنگ ورودی‌های شبکه را نشان می‌دهند و فلش‌های قرمز رنگ مقایسه خروجی شبکه با خروجی مورد انتظار را.

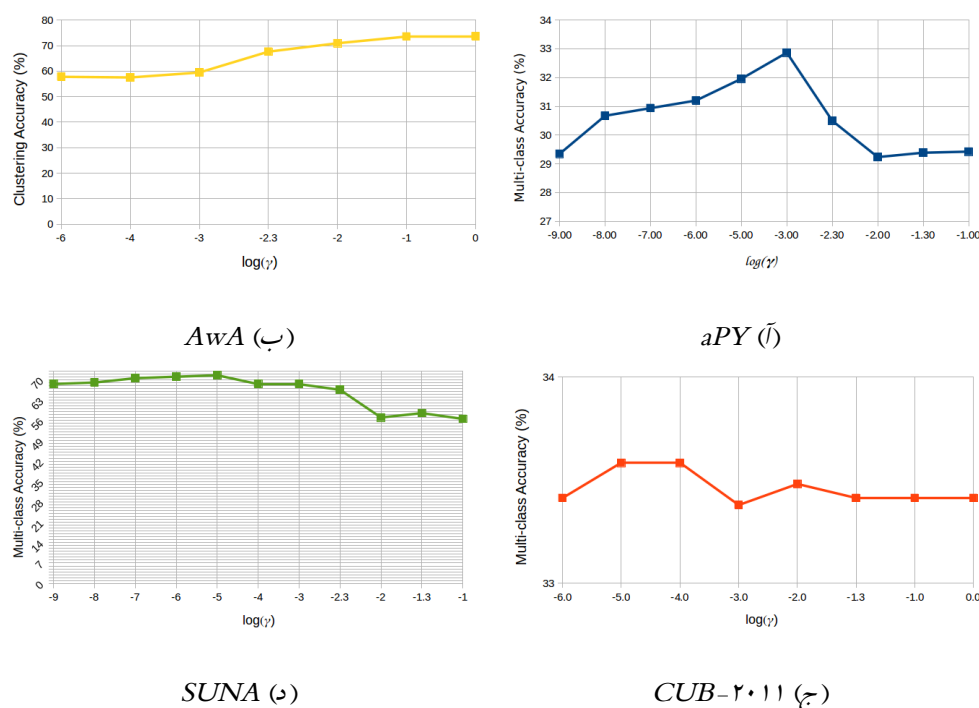


شکل ۳-۴: نمایش دوبعدی بوسیله t -SNE برای ده دسته‌ی آزمون از مجموعه دادگان AWA با ده رنگ متفاوت نشان داده شده است. درستی فرض قابل خوشه‌بندی در تصویر مشخص است، یعنی ویژگی‌های استخراج شده با استفاده از شبکه‌های ژرف توانایی ایجاد تمایز بالا میان دسته‌ها را دارا هستند و نمونه‌های هر دسته نیز نزدیک به یکدیگر هستند.

توابع مطابقت پیشین، که میزان شباهت هر نمونه را به طور جداگانه با توصیف دسته‌ها محاسبه می‌کردند، متفاوت است و همه‌ی نمونه‌ها در تعیین برچسب یکدیگر موثر هستند. در این حالت هر خوشه باید یک برچسب دریافت کند و برچسب اختصاص یافته به هر خوشه، توسط تمام اعضای آن به ارث برده می‌شود. این تابع مطابقت تا کنون در روش‌های موجود برای یادگیری صفرضرب استفاده نشده است. نسخه‌های متفاوتی از این تابع مطابقت، بر حسب چگونگی تعیین برچسب هر خوشه، قابل ارائه است. ما در اینجا دو مورد از آن‌ها را بیان می‌کنیم. یک شیوه برای انتساب برچسب به خوشه‌ها، استفاده از رای اکثریت است؛ در این حالت بایست ابتدا یک پیش‌بینی برای همه نمونه‌های آزمون صورت بگیرد (برای مثال با استفاده از روش معرفی شده در بخش ۳-۲)، فرض کنید مقادیر این پیش‌بینی را با z_n برای $N_s < n \leq N_s + N_u$ نشان دهیم. همچنین یک خوشه‌بندی روی داده‌ها انجام شده که آن را با r_n برای $N_s < n \leq N_s + N_u$ نشان می‌دهیم. حال $\ell(k)$ که برچسب خوشه‌ی k -م است از رابطه زیر تعیین خواهد شد:

$$\ell(k) = \arg \max_{n_s < i \leq n_s + n_u} \left[\sum_{m=N_s+1}^{N_s+N_u} \mathbb{1}(r_m = k) \times \mathbb{1}(z_m = i) \right]. \quad (14-3)$$

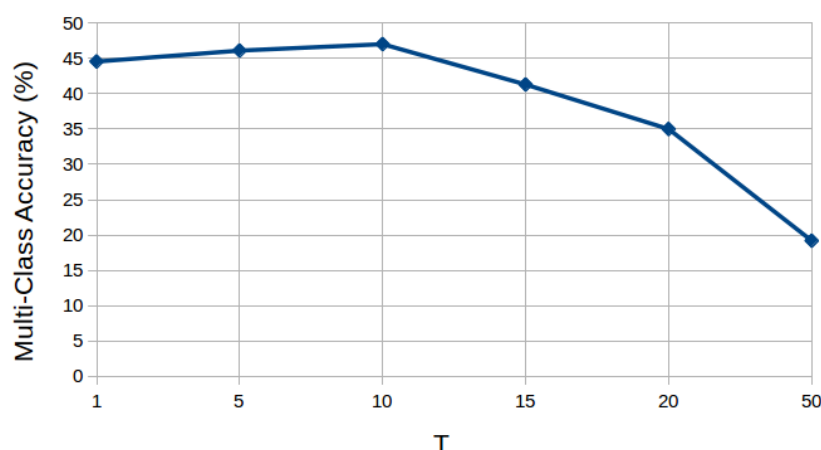
این نسخه از تابع مطابقت پیشنهادی قابل اضافه شدن به روش‌های دیگر نیز هست. به این صورت که پیش‌بینی‌های انجام



شکل ۴-۱: میزان دقت دسته‌بندی چند دسته‌ای در شبکه چندوظیفه‌ای ارائه شده (نسخه یک لایه) بر حسب \log_1 پارامتر γ در معادله (۲-۳).

۴-۵ بررسی خوشه‌بندی نیمه‌نظارتی

در این بخش به بررسی عملکرد روش خوشه‌بندی نیمه‌نظارتی ارائه شده در بخش ۳-۵ می‌پردازیم. برای این منظور روش ارائه شده را روی هر مجموعه داده اجرا کرده، خوشه‌های مربوط به دسته‌های دیده‌شده را کنار گذاشته و هر یک از خوشه‌های دیگر را به یک دسته از دسته‌های آزمون نسبت می‌دهیم. برای این کار در هر خوشه بر اساس برجسب صحیح نمونه‌ها رای‌گیری می‌شود و برجسبی که بیشتر اعضای آن خوشه آن را دارا هستند به کل اعضای خوشه نسبت داده می‌شود. نتیجه با برجسب‌های صحیح مقایسه شده و دقت دسته‌بندی چنددسته‌ای در جدول ۴-۴ گزارش شده است. برای مقایسه عمل‌کرد، آزمایش مشابهی را با روش k-means اجرا می‌کنیم. به این صورت که الگوریتم k-means را با $k = n_s + 2n_u$ اجرا کرده و با هر خوشه با رای‌گیری برجسب یکی از دسته‌های دیده نشده را نسبت می‌دهیم. نتایج مربوط به این آزمایش نیز در جدول ۴-۴ گزارش شده است.



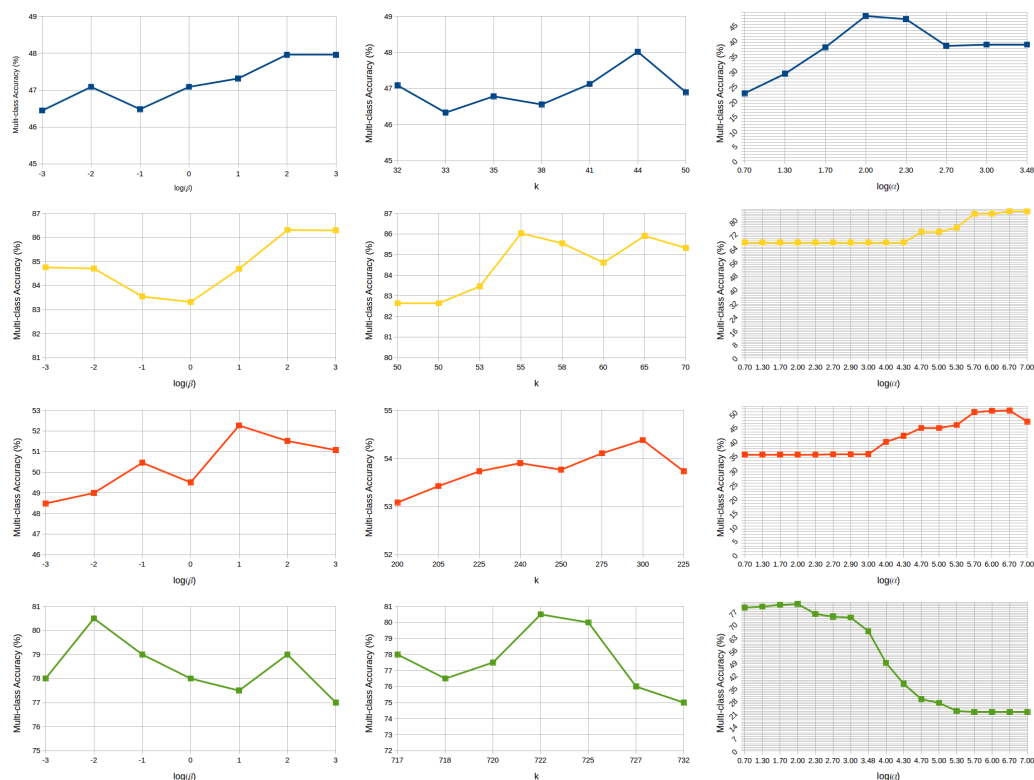
شکل ۴-۲: بررسی میزان دقت دسته‌بندی بر حسب پارامتر T در رابطه (۳-۱۱) برای مجموعه داده‌های *aPascal/aYahoo*: افزایش T در ابتدا می‌تواند باعث افزایش دقت شود ولی ادامه افزایش آن باعث نزدیک شدن مقادیر هیستوگرام به یکدیگر و کاهش دقت دسته‌بندی می‌شود.

در تمامی آزمایش‌ها کمتر از ۵ دقیقه بوده است.

نتایج مربوط به این روش در جدول ۴-۵ با عنوان نگاشت به هیستوگرام آمده است. همان‌گونه که مشاهده می‌شود این روش با اینکه از روند ساده و همچنین سریعی بخاطر استفاده از الگوریتم‌های بهینه‌سازی تصادفی برخوردار است، به نتایج بهتری نسبت به روش‌های پیشین دست یافته است و تنها از روش بسیار اخیر ارائه شده در [۳۷] دقت کمتری داشته است. این در حالی‌ست که در سایر روش‌های مبتنی بر هیستوگرام ([۳۶، ۳۷]) از روندهای بهینه‌سازی همراه با محدودیت استفاده می‌شود که بسیار کندتر هستند. برای مثال حداکثر زمان اجرا در [۳۶] روی چهار مجموعه داده‌های مورد بررسی ۳۰ دقیقه اعلام شده است در حالی‌که در آزمایشات انجام شده زمان آموزش شبکه پیشنهادی کمتر از ۵ دقیقه بوده است. همچنین به علت محدودیت‌های روش‌های بهینه‌سازی محدب، این روش‌ها در مجموعه داده‌های بزرگ مانند ImageNet قابل استفاده نیستند در حالی‌که روش پیشنهادی دارای قابلیت مقیاس‌پذیری و استفاده در مجموعه داده‌های بزرگتر است.

۴-۷ دسته‌بندی با روش خوشه‌بندی و یادگیری نگاشت مجزا

در این بخش به بررسی عملی روش پیشنهادی برای روش خوشه‌بندی و یادگیری نگاشت مجزای نیمه‌نظارتی می‌پردازیم که در بخش ۳-۶ معرفی شد و مراحل آن در الگوریتم ۲ ذکر شده است. این روش مبتنی بر یک خوشه‌بندی روی داده‌های



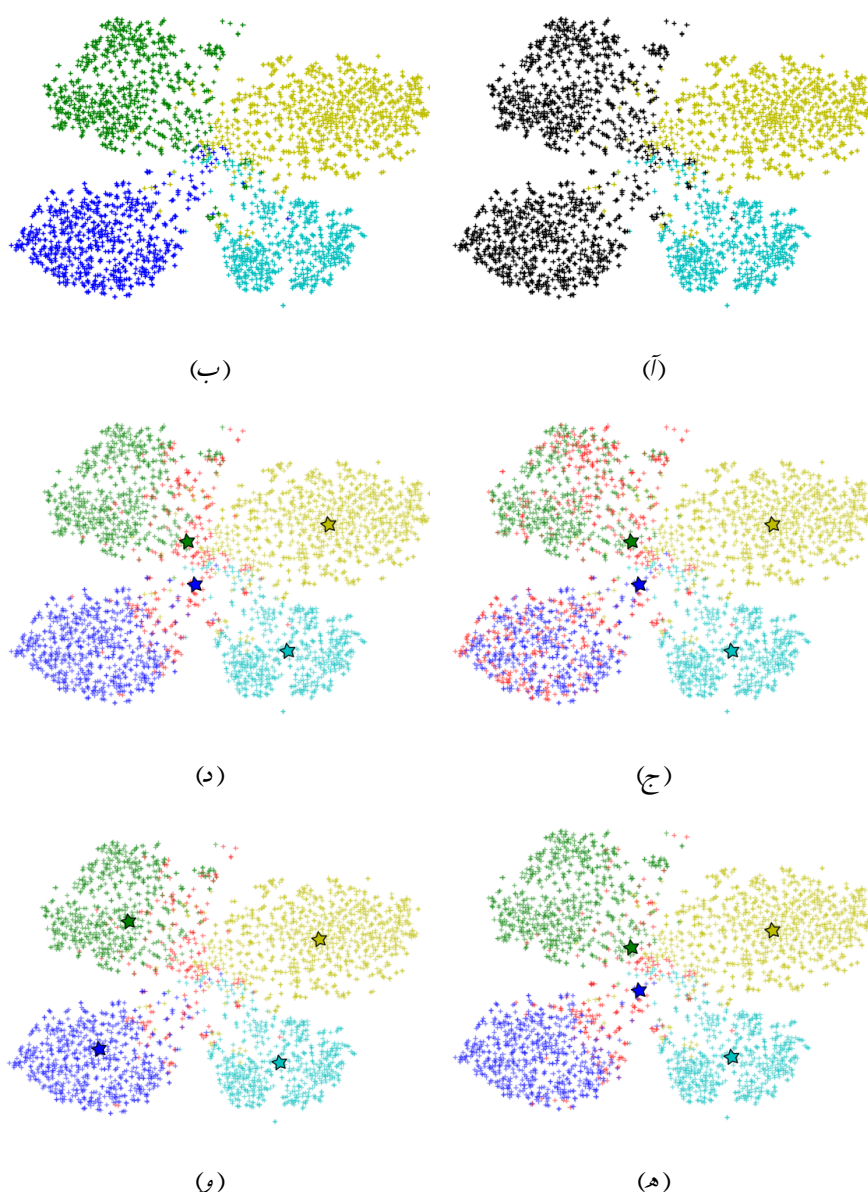
شکل ۴-۳: تاثیر پارامترهای روش خوشه‌بندی و یادگیری نگاشت مجزای نیمه‌نظارتی. سمت چپ: نتیجه دقت دسته‌بندی چند دسته‌ای بدست آمده بر حسب پارامتر α در رابطه (۳-۲۳) که اهمیت جمله منظم‌سازی را نشان می‌دهد. همان‌طور که مشاهده می‌شود، عمل‌کرد روش به این پارامتر حساس است. وسط: نتیجه دقت دسته‌بندی چند دسته‌ای بدست آمده بر حسب تعداد خوشه‌ها در خوشه‌بندی نیمه‌نظارتی. با توجه مقیاس این نمودار مشخص می‌شود که دقت حاصل شده حساسیت کمی نسبت به این پارامتر دارد. سمت راست: نتیجه دقت دسته‌بندی چند دسته‌ای بر حسب پارامتر β در خوشه‌بندی نیمه‌نظارتی (رابطه (۳-۱۷)).

برای راحتی مقایسه محور عمودی همه‌ی نمودارها با بازه‌های یک درصدی تقسیم‌بندی شده‌اند.

سطر اول (آبی‌رنگ): مجموعه دادگان aPY. سطر دوم (زرد رنگ): مجموعه دادگان Awa. سطر سوم (قرمز رنگ): مجموعه دادگان CUB-2011. سطر چهارم (سبز رنگ): مجموعه دادگان SUNA.

R به صورت بیان شده در الگوریتم ۳ به طور متوسط $6/8\%$ دقت بالاتری در دسته‌بندی نسبت به مقداردهی D با رابطه (۳-۲۳) دارد. دلیل این موضوع همان‌طور که در بخش ۳-۷ بیان شد استفاده از اطلاعات بدون نظارت نمونه‌های آزمون در بدست آوردن مقدار اولیه برای R است در حالیکه در مقداردهی اولیه D تنها نمونه‌های آموزش دخالت دارند.

به علت حساسیت نتایج این روش به پارامترهای آن (مقادیر λ و η در رابطه (۳-۲۵))، مقادیر آن‌ها توسط روند اعتبارسنجی شرح داده شده در بخش ۴-۲ تنظیم می‌شود.



شکل ۴-۴: نمایش دوبعدی چهار دسته از مجموعه دادگان AWA با استفاده از نگاشت t -SNE، دو دسته‌ی دیده شده شامل بزرگوزن (فیروزه‌ای) خرس گریزلی (زرد) و دو دسته‌ی دیده نشده شامپانزه (آبی) و پاندا (سبز). تصاویر با نماد بعلاوه و نگاشت توصیف دسته‌ها در فضای تصاویر با ستاره نشان داده شده است. در تصاویر (ب) تا (و) نقطه‌های قرمز نمونه‌هایی که را نشان می‌دهد که دسته‌ای به جز چهار دسته‌ی موجود در شکل برای آن‌ها پیش‌بینی شده است. (آ) دسته‌های دیده شده با برجسب صحیح و دیده‌نشده با رنگ مشکی (ب) نمایش برجسب صحیح برای تمامی دسته‌ها (ج) توصیف‌ها با نگاشت (۳-۲۳) به فضای تصاویر برده شده‌اند و دسته‌بندی با دسته‌بند نزدیک‌ترین همسایه انجام شده است. (د) نگاشت مانند حالت قبل و دسته‌بندی با تابع مطابقت پیشنهادی به همراه خوشه‌بند k -means (ه) نگاشت مانند حالت قبل و دسته‌بندی با تابع مطابقت پیشنهادی به همراه خوشه‌بند نیمه‌نظارتی پیشنهاد شده (و) دسته‌بندی و نگاشت با استفاده از روش پیشنهادی برای یادگیری نگاشت و خوشه‌بندی توأم.