

## رسیدن به شباهت معنایی تصویر از طریق دسته بندی حاصل از جمع سپاری

### چکیده

تعیین شباهت بین تصاویر از جمله دسته بندی تصویر، برچسب زنی تصویر و همچنین بازیابی، مرحله اساسی در بسیاری از برنامه های کاربردی است. روش های اتوماتیک برای ارزیابی شباهت اغلب ناکارآمد است و وقتی که زمینه معنایی برای کار نیاز است نیاز به قضاوت انسان پیش می آید چنین قضاوتهایی را می توان از طریق فنون جمع سپاری بر اساس کارهای ارائه شده توسط کاربران وب جمع آوری نمود. با این حال برای ممکن ساختن برآورد شباهت تصاویر با زمان و هزینه معقول ایجاد کارها برای مقدار انبوه باید به طریقه دقیقی انجام شود. ما مشاهده کردیم که فواصل بین همسایگی های محلی اطلاعات ارزشمندی فراهم می کند که به ما این امکان را می دهد که معیار شباهت کلی را سریع و دقیق ایجاد کنیم. این ملاحظه کلیدی ما را به سمت راه حلی براساس وظائف دسته بندی و مقایسه تصاویر نسبتاً مشابه هدایت می کند در هر جستجو، اعضای گروه مجموعه کوچکی از تصاویر را در داخل صندوق های خود جمع آوری می کنند. نتایج حاکی از شباهت های نسبی زیاد بین عکس ها می باشد، که برای ایجاد معیار شباهت مورد استفاده قرار گرفته اند. این معیار به صورت تصاعدی اصلاح می شود و برای ترتیب دادن تحقیقات بهتر و مکانی تر در تکرارهای بعدی بهبود می یابد. ما موثر بودن روش خودمان را بر روی مجموعه داده ها وقتی زمینه میدانی وجود داشته باشد و روی مجموعه ای از تصاویر در صورتی که شباهتهای معنایی را نمی توان اندازه گیری کرد اثبات می کنیم. در حالت خاص نشان می دهیم روش ما گزینه بهتری از روش های مرجع است و موثر بودن تحقیقات جمع آوری و فرآیند اصلاح ما را بهبود می بخشد.

**کلمات کلیدی:** جمع سپاری، شباهت تصویر، معیار فاصله تصویر

## 1. مقدمه

درسالهای اخیر پیشرفت های بسیاری در ضبط عکس در ابزارهای همراه حاصل شده است که کاربران نهائی را تشویق می کند عکس های بیشتری با کیفیت بالاتر ضبط نمایند. در نتیجه این امر در مجموعه های عکس، هم در رایانه های شخصی و هم در وب سایت هائی نظیر فیسبوک، فلیکرواینستاگرام و فوردائمی وجود دارد. چنین مجموعه های بی شماری نیاز به روش هائی برای دسته بندی عکس ها ، برچسب زنی تصویر و به ویژه ذخیره عکس دارد که به کاربران این اجازه را می دهد به سرعت عکس مناسب برای نیازشان را بیابند. این روش ها ضرورتاً متکی بر موجود بودن شباهت های دودو بین دو عکس در مجموعه هستند.

بسیار دشوار است که یک معیار فاصله ای تعیین کنیم تا شباهت شناخت حضوری یا معنایی بین تصاویر را به خوبی ضبط کند. روش های تحلیلی برای محاسبه چنین معیار زمانی که شباهت ها از زمینه معنایی بر می آید ناکارآمد است . این ناکارآمدی ممکن است شامل روابط فرآر نظیر هیجان و احساس مشابه که توسط تصویر القا می شود. (مثلاً تصایری که ترس یا آرامش را انتقال می دهند) تصاویر اشیائی که از لحاظ معنایی مرتبط هستند (مثلاً انواع مختلف وسائل باغ) ، همانندی بین افرادی که عکسشان گرفته شده و غیره . به عنوان نمونه به شباهت بین تصاویر پوستر فیلم در شکل یک دقت کنید . شناسائی چنین شباهتهائی معمولاً توسط انسان به آسانی انجام می شود ، اما با این حال همراه باعث بروز ایراد مشکل شدید محاسباتی خواهد بود.

بنابر این راه حل علمی جمع آوری اطلاعات درباره شباهت های معنایی بین تصاویر افراد است ، برای مثال استفاده از تکنیک جمع سپاری<sup>1</sup> . این روند در کار اخیر برای جمع آوری معیارهای نوع شباهت اثر بخش بوده است. [9,13] . کارنوعی مقایسه که گروه انجام می دهد از فرم های زیرتبعیت می کند: سه تصویر A ، B و C داده شده است ، انتخاب کنید که A بیشتر شبیه B است تا C ( پرسش سه گانه) . با فرض پاسخهای درست پرسش ، پرسش درباره هرتصویر ، پرسش سه گانه نتیجه معیار شباهت نسبی در کل مجموعه تصاویر را در پی خواهد داشت . با این حال

---

<sup>1</sup>جمع سپاری نام عمومی پردازش هائی است که شامل ارائه کارهائی با مقیاس کوچک برای گروه استفاده کننده از وب و تلفیق جواب های گروه برای رسیدن به هدف با مقیاس وسیع نظیر ایجاد بانک اطلاعاتی بزرگ





شکل 2. بدون زمینه ، شباهت بین تصاویر مبهم خواهند بود



شکل 3. با زمینه ، قابل مشاهده است که تصاویر متعلق به دو مکان مشخص است ، لندن و پاریس

مشاهده کلیدی این کار آن است که معیار شباهت را می توان به صورت اثر بخش تری از تصاویر شبیه به هم نسبت به تصاویری که شباهتی به هم ندارند تدوین کرد. این به ویژه در زمینه شباهتهای معنایی و در صورتی که شباهتهای مکانی بیشتر معنادار است حقیقت دارد. با پیگیری این مشاهدات ، ما الگوریتم توافقی جدیدی ایجاد کردیم که منظور از آن تدوین پرسشنامه هائی است که حتی الامکان از یک محل باشند . در اینجا چالش این است که شباهت ها از قبل معلوم نیستند . بنابر این الگوریتم ما به صورت تکراری کار می کند. در هر مرحله مابرای گروه پرسش ایجاد کرده و ارائه می دهیم. بعد از جمع آوری اطلاعات به تدریج پرسش ها را اصلاح می کنیم تا روی تصاویر مشابه در همسایگی های نزدیکتر محلی تمرکز کنیم . مقایسات محلی شباهت در فضای اقلیدسی جای داده می شوند تا برآورد اصلاح شده برای معیار کلی بدست آید. سپس این معیار اصلاح شده در مرحله بعدی برای محاسبه پرسش نامه هائی که بیشتر روی مکان تاکید دارند به عنوان ابزار مورد استفاده قرار می گیرند. این روش پیش رونده برآورد شباهت معنادار را همگرا می سازد.

**ارزیابی و بررسی آزمایشی** برای سنجش کارآمد بودن روش ما تکنیک را هم با داده های ساختگی و داده های واقعی گروه در سیستم نمونه به کار بستیم . اول ما تکنیک خود را بر روی دو مجموعه داده از تصاویر در حالتی که معلوم بود آزمودیم ، نتایج را آزمایش کردیم و با روش پایه که از تعداد مساوی پرسش نامه با محل واقعی معلوم استفاده می کرد اما به صورت تصادفی انتخاب شده بودند مقایسه کردیم . دوم ، ما تصاویر k-NN را برای مجموعه

داده های تصاویر زمینه میدانی درحالتی که محل واقعی معلوم نبودنتایج را به صورت دستی ارزیابی کردیم . در آخر ما اثر پارامترهائی نظیر تعداد مراحل و پرسش نامه ها را در مجموعه آزمایشات ساختگی بررسی کردیم .نتایج آزمایشی ما کارآمدی روش ما را برای محاسبه شباهت معنایی تصویر که منحصرأ بر اساس پاسخ های گروه است را ، بهبود می بخشد و این در حالی است که از تعداد کمی از پرسشنامه دسته بندی استفاده شده است .

## 2. کار مرتبط

دسته بندی تصاویر مسئله ای که خیلی خوب مورد بررسی قرار گرفته است . الگوی عام بر اساس توصیفگرهای تصویر نظیر هیستوگرام رنگ تصاویر ، توصیفگر ها بر اساس SIFT یا توصیفگرهای GIST است . فاصله بین دو تصویر به عنوان فاصله اقلیدسی بین توصیفگرهای تصویر تعریف شده است ، دراولویت تمامی تکنیک های آموزشی که در ماشین می تواند برای یافتن شباهتها به کارگرفته شود تا شباهتها یا مجموعه عکس ها یافته شود. با این حال چنین متدهائی از روش (BoF) (حالت های کیفی پر از حالتها ) اجزا بصری استفاده می کنند که یا به تصویر به صورت دستی چسبانده شده یا از بافت متنی یک صفحه وب استفاده می کنند. یا این حال چنین متدهائی وقتی دسته بندی مبتنی بر حالتهای معنایی غنی باشد و ممکن است تنها بخشی از آن در برچسب منعکس شده باشد ، ناکارآمد است . برای مثال تصاویر پل های لندن و پاریس حاکی از بسیاری حالات معنایی نظیر استایل ، مواد ساختمانی و محیط عمومی هستند. تصاویر و برچسب های شرح دهنده آنها نه تمام این حالات و نه اهمیت نسبی آنها را برای تعیین شباهت ضبط نمی کنند .

مشکل نبود حالتهای معنایی را می توان با روش های یادگیری که تا بخشی نظارتی هستند و متکی بر برچسب زنی دستی مجموعه کوچکی از عکس های دوبدو و سه گانه هستند بیشتر کاهش داد تا اینکه هر تصویر از کل مجموعه برچسب گذاری شود . حجم وسیعی از کار تلاش برای کلاس بندی تصاویر با استفاده از چنین متدهائی است ، معمولاً با برچسب زنی دوبدو شامل محدودیت های همسانی ( و گاهی اوقات ناهمسانی) مثلاً آیا دوگانه متعلق به یک کلاس است یا نه.[1,2,19,21] محدودیت های سه سویه بیشتر مربوط مقایسات نسبی تصاویر است چرا که

آنها فاصله دوجفت عکس را مقایسه می کنند.[6,9,11,13,16] سپس محدودیت ها را می توان برای آموزش یک معیارفاصله بین عکسها به کارگرفت. به ویژه کار بر روی انتخاب تطبیقی سه گانه های مناسب بر پایه داده های گروه تمرکز یافت [6]. در کار اخیر [9,11,13] مقایسات سه سویه از بین اعضای گروه جمع آوری شده تا درباره شباهت های استایل آگاهی کسب شود استفاده از مقایسات سه گانه دارای این عیب بود که در کار با آن مواجه شدیم : این روش نیازمند انجام کارزیاد توسط گروه بود و به کاربران زمینه لازم داده نشده بود. این عیوب همچنین دربررسی تمرکز یافته [20] روی طراحی مجدد رابط برای انجام مقایسات بیشتر از کار افراد گروه مورد توجه قرار گرفت این کار با درخواست از کاربران برای انتخاب تصایر شبیه تر  $X$  به تصویر داده شده غیر از تصاویر  $Y$  انجام شد. رابط های جدید [20] یک قدم جلوتر از سه گانه هستند اما در مقایسه با کار ما بررسی آنها به نکته چگونگی انتخاب موثر تصاویر برای مقایسه توجه نشده است .

کار دیگری که بسیار مرتبط به کار دسته بندی توسط گروه ما است [7]، به دسته بندی تصویر توسط گروه توجه داشته است. هر عضو گروه نمونه ای از چند تصویر دریافت می کند (یک پرسش) و آنها را در داخل گروه های کلاس بندی می کند. این ورودی مدل بیزی روش های ممکن مورد استفاده توسط برای کلاس بندی هر تصویر را برآورد می کند. این کار مشابه ما به کاربر این امکان را می دهد تا تا هر تصویر مجموعه کوچکی از تصاویر را کلاس بندی کند ، و نیز در ایده بهبود نتایج دسته بندی با به کارگیری مجدد تکنیک روی دسته های بدست آمده شبیه کار ما می باشد. با این حال تکنیک آنها برای محاسبه شباهتهای تصاویر طراحی نشده است . در مقابل ما بهبود مستمر را برای تعیین شباهتهای تصویر با همرسی سریعتر به کار بستیم . ما عملکرد تکنیک های خود را با [7] دربخش 4 مقایسه نمودیم.

کار [22] توصیه می کند تنها پاسخ پرسش ها را دربخش کوچکی از داده دریافت کنیم و از تکنیک های پایان یابی ماتریس اختصاصی برای تکمیل کلاس بندی های نامعلوم استفاده کنیم تا اینکه نیار باشد که هر تصویر در یک پرسش به گونه [7] وجود داشته باشد. این کار متعامد بر کارهای ماست و می توان آن را به کارگرفت اگر، تعداد پرسش هائی که می توان پرسید متناسب با تعداد تصاویر باشد.

جمع سپاری برای کارهای مرتبط با کارما نظیر جور کردن رکوردها روی تصاویر [10] گروه بندی و top-k [5] و جور کردن کلیت مورد استفاده قرار گرفته است. با این وجود نه در هیچکدام از کارهای قبلی به مسئله آموزش یک معیار شباهت تصویر توجه شده و نه می توان آنها را به صورت مستقیم در این کار، به کار بست. برای مثال k-NN (k-Nearest Neighbors algorithm) به منظور یافتن top-k شبیه ترین تصاویر برای هر تصویر مورد ملاحظه قرار گیرد؛ با این حال به کاربرد روش [5] به صورت مجزا برای هر تصویر ناکارآمد است .

### 3. الگوریتم

ما بعداً متد ایجاد پرسش ها را برای گروه براساس معیار برآورد شده شباهت و بهبود معیار بر اساس پاسخ های اخذ شده از گروه شرح خواهیم داد . قصد ما استفاده از پرسش هائی است که مشتمل بر تصاویری از یک محل همسایگی است که برای تعیین معیار کلی موثرتر است .

الگوریتم ما پرسش های دسته بندی را با استفاده از تصاویر  $nq$  ایجاد می کند . پاسخ دریافت شده از اعضای گروه برای قرار دادن تصویر دسته های  $nc$  است . گروه منبع نسبتاً گرانی به لحاظ نایابی ، تلاش انسان و هزینه پولی است . بنابر این در بسیاری از موارد تجربی ، تعداد کل پرسشهایی که می توان پرسید به لحاظ بودجه از پیش تعیین شده محدود است . با چنین بودجه مشخصی هدف ، الگوریتمی که ما بوجود آوردیم بهره گیری از پرسش ها به بهترین وجه ممکن با منحصراً با در نظر گرفتن همسایگی های محلی است. حاصل این امر فرآیند تکراری است درحالی که همسایگی ها مطابق با نتایج پرسش ها تغییر می یابند.

متد ما فواصل محلی را با نگهداشتن دسته بندی از کل مجموعه در فضای اقلیدسی برآورد می کند که در آن فواصل محاسبه می شوند . دسته بندی به صورت تصادفی آغاز می شود و همسایگی های محلی به صورت مستمر بهبود می یابند. دسته بندی این اطمینان را می دهد که فواصلی که حتی پرسیده نشده اند در اطلاعات ناکاملی که از فواصل پرسش شده سازگار باشد . برای بهبود دسته بندی همسایگی های محلی، ما پرسش هائی در دسته های کوچک به کاربران ارائه دادیم و بعد از ارائه هر دسته از پرسش ها دسته بندی را بهبود بخشیدیم. به صورت جالب توجهی

پرسش کردن همسایگی های محلی از دسته بندی به صورت اثر بخشی بهبود یافت حتی در مراحل آغازین که تصاویر ضرورتاً از لحاظ معنایی نزدیک نبودند ، چون چنین پرسش هائی در همان همسایگی محدودیت های زیادی ایجاد می کردند. به علاوه ما می خواستیم در هر تکراری همسایگان نزدیکی که تا بحال از لحاظ معنایی شبیه بوده اند را حفظ کنیم . حتی در دسته بندی تصادفی ، پرسش ها بر مبنای همسایگی به تشخیص و حفظ موارد کمک می کند اگر همسایگان نیز از لحاظ معنای شبیه باشند .

مراحل اصلی الگوریتم در الگوریتم 1 نمایش داده شده است : به عنوان ورودی الگوریتم تعداد کل پرسش های مجاز را (budget) دریافت می گیرد و تعداد پرسش ها در هر تکرار (batch\_size) . نتایج پرسش ها در دسته بندی (E) و معیار کلی فاصله ایجاد شده (D) تجمیع شد. خروجی الگوریتم معیار فاصله محاسبه شده بر اساس آخرین و بهبود یافته ترین دسته بندی است .

---

**Algorithm 1: CrowdSter(budget, batch\_size)**

---

```

1: E = EmbedData()      // random embedding
2: num_of_queries = 0
3: while num_of_queries < budget do
4:   Q = SelectQueries(E, batch_size)
5:   R = RunQueries(Q)  // using the crowd
6:   D = DistanceFromEmbedding(E)
7:   D = UpdateDistances(D, R)
8:   E = EmbedData(D)
9:   num_of_queries += batch_size
10: end while
11: D = DistanceFromEmbedding(E)
12: Output D

```

---

در پرسش دسته بندی برای مجموعه تصاویر I ، ما پرسشی Q تعریف کردیم به عنوان زیر مجموعه I حاوی تصاویر  $nq$

پاسخ به هر پرسش بخشی است از Q به داخل دسته های گسسته  $C_1, \dots, C_{nc} \subseteq Q$  . از این پاسخ ما مقایسات شباهتها را استخراج کردیم : تصاویر داده شده  $x, y$  در در دسته  $C_i$  و تصویر سوم  $z$  در دسته متفاوت  $C_j$  ، ما نتیجه گیری کردیم که  $\Delta(x, y) < \Delta(x, z)$  اگر  $\Delta$  نشان دهنده معیار شباهت باشد. همان قدر که  $nq$  افزایش می یابد، مقایسات بیشتری اخذ می کنیم اما تعداد تصاویر در یک پرسش باید تا اندازه ای کوچک باشد تا اعضای گروه بتوانند آنها را مشاهده نمایند [10] . در آزمایشات ما مشخص شد که تعداد 20 برای  $nq$  میزان خوبی برای پایاپای بودن



سادگی و اثر بخش بودن است. همسو با این دریافتهیم که تنظیم تعداد دسته های  $nc$  بهینه است که بین حصول مقایسات بیشتر (مقادیر کمتر  $nc$ ) و حذف سریع تصاویر مشابه (مقادیر بیشتر  $nc$ ) توازن ایجاد می کند .

ایجاد پرسش ها در الگوریتم ما پرسش ها بر اساس حذف از فازهای قبلی ایجاد شده اند. در هر فاز ، پرسش هائی (a) ایجاد کردیم که محلی هستند و (b) مجموعه عکس هائی را پوشش می دهد که حتی الامکان دقیق باشند. برای چنین کاری ما تصاویر تصادفی به صورت یکنواخت (یکدست) نمونه گرفتیم در حالی که مطمئن شدیم نزدیکترین همسایگان هم نیستند. وقتی چنین نمونه هائی باقی ماندند دوباره شروع کردیم . برای هر عکس نمونه گرفته شده  $k$  آن را ، نزدیکترین همسایه در دسته بندی ، یافتیم . سپس خارج از این همسایه ها ما زیر مجموعه تصادفی اندازه  $nq$  نمونه گرفتیم و در پرسش بعدی مورد استفاده قرار دادیم .

دسته بندی ما دسته بندی تمام تصاویر را، که به تدریج با هر دسته از پرسش ها بهبود یافته بود ، نگهداری کردیم. دسته بندی حاوی اطلاع دقیقی در مورد فاصله سازگار بین هر دو گانه از تصاویر است که در فاز بعدی مورد استفاده قرار گرفته است. قبل از اینکه اولین پرسش به کاربران ارسال شود تصاویر با استفاده از پراکندگی تصادفی همسان در فضای اقلیدسی دسته بندی شده است . برای بهبود تدریجی دسته بندی ما فاصله بین هر جفت از عکس ها را محاسبه کردیم و فواصل را برحسب نتایج نتایج پرسش بهنگام سازی نمودیم و تصاویر را مجدداً با استفاده از این فواصل بهنگام سازی شده دسته بندی نمودیم . این کار فواصل بهنگام سازی شده را همسان نمود و هر ناهمسانی بین آنها را برطرف نمود. برای محاسبه دسته بندی به عنوان ورودی فاصله بین هر جفت از عکسها را به مقیاس دهی چند بعدی (MDS) دادیم و محاسبه انجام شد.

به صورت ویژه تر می خواهیم دسته بندی بیابیم که تنها اهمیت لحاظ شده در آن فواصلی باشد که درباره آنها اطلاعات داریم (از طریق نتایج پرسش)، و از تمام فواصل دیگر چشم پوشی شده باشد. برای این کار از الگوریتم **Sammon Projection** [14] استفاده شد که یک تکنیک مقیاس بندی چند بعدی است و دسته بندی را با استفاده از تابع استرس و شیب نزولی تعریف می کند. تابع سنگین استرس برای اهمیت دادن به فواصل مرتبط و صرف نظر از سایر فواصل با دادن وزن خیلی کم به آنها، به کار گرفته می شود . تمام وزن ها با مقدار خیلی کم  $\epsilon$

آغاز می شود. در هر مرحله ما برای هر فاصله بهنگام سازی شده  $1$  را اختصاص می دهیم. به فواصلی که در مراحل پیشین بهنگام سازی شده اند مقدار  $1$  تخصیص داده می شود بنابراین یک بار که یک بار دوگانه عکس ها مورد پرسش قرار میگیرند وزن آن در مراحل بعدی نظر گرفته می شود.

بهنگام سازی فاصله برای بهنگام سازی فاصله، تمام نتایج پرسشها در دسته تجمیع شده و تحلیل می شود. برای هر دوگانه تصویر در هر پرسش ما به نتیجه یک پرسش را مثبت تلقی می کنیم اگر تصویر به همان دسته تخصیص یافته باشد و به صورت منفی تلقی می کنیم اگر به دسته های متفاوت تخصیص داده شده باشد. نتایج پرسش درباره دوگانه تصویر اگر نتایج مثبت نسبت به نتایج منفی بیشتری باشد، کوتاه ترمی شود و بلندتر می شود اگر دوگانه تصویر دارای نتایج پرسش منفی بیشتری باشد.

فواصل بین بین تصاویر وقتی مساوی باشد و فواصل بین دوگانه های تصویر که در همان پرسش حضور نداشته باشند در نتایج تاثیر داده نمی شوند. فواصل با تقسیم  $\beta$  کوتاه شده و با ضرب در  $\beta$  درازتر می شوند. در آزمایشات ما  $\beta$  تنظیم شده است تا توجه شود که ما نمی توانیم تعداد دفعاتی که دوگانه تصاویر در یک دسته پرسش قرار گرفته اند را به حساب آوریم. به عنوان مثال دوگانه تصویری که دارای دوتیجه مثبت پرسش از دو باشد به همان طریق به صورت دوگانه تصویری که دارای سه مثبت از چهار مثبت نتیجه پرسشها است بهنگام سازی می شود. چون مراحل کوتاه در نظر گرفته شده، احتمال اینکه دوگانه تصویر مشابه در بسیاری از پرسش ظاهر شود بسیار کم است و اطلاع از نسبت دقیق بین نتایج مثبت و منفی برای تصادفی بودن بیش از حد دارای حساسیت است.

#### 4. آزمایشات

برای ارزیابی کارآمدی روش ما سه مجموعه آزمایشات را ترتیب دادیم که در زیر شرح داده شده است. اول برای سنجش درستی روشمان مجموعه آزمایشاتی در مقیاس کوچک را روی مجموعه ای از داده ها ترتیب دادیم وقتی زمینه میدانی شناخته شده باشد. این زمینه میدانی امکان ارزیابی کیفیت نتیجه را ایجاد می کند. دوم عملی بودن روش را برای تصاویری غنی لحاظ معنایی با استفاده از مجموعه های گسترده تر تصاویر با زمینه میدانی نامعلوم

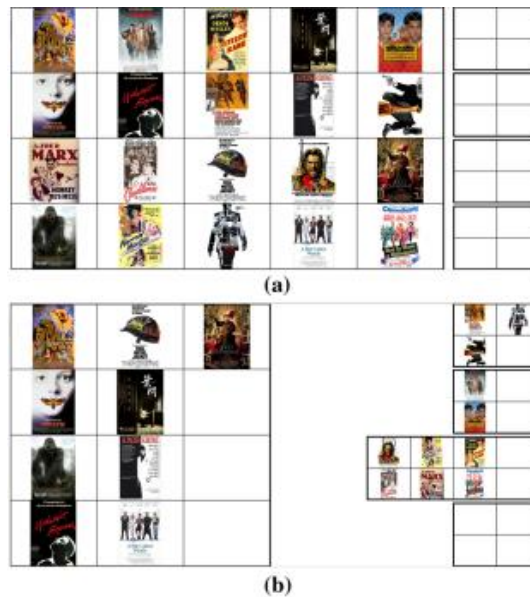
آزمودیم . درانتها برای بررسی بیشتر هر بخش از راه حل، آزمایشات ساختگی ترتیب دادیم(درحالت زمینه میدانی معلوم و درحالتی که به جای پاسخ گروه واقعی به پرسش ها ، شبیه سازی گروه صورت گرفته بود). ما پارامترهای مختلفی از سیستم را تغییر دادیم و نتیجه آن را روی کیفیت خروجی مشاهده نمودیم . در تمام مجموعه آزمایشات ما بیشتر نتایج بدست آمده را با الگوریتم های پایه جایگزین مقایسه کردیم .

• تصادف انتخاب تصادفی پرسش ها معادل با اجرای الگوریتم ما در یک مرحله

• دسته گروه با استفاده از متد [7]. نتایج این متد برای شناسایی مجموعه ها هدف گذاری شده بود اما برای محل فضائی واقعی برای هر تصویر که از آن به عنوان جایگزین جای سپاری ما استفاده شده بود نیز شامل شد.

• بر مبنای حالت برآورد شباهت تصاویر بر مبنای حالت های تصویر که به صورت خودکار استخراج شده بودند و به عنوان مبنا در نبود زمینه میدانی عمل می کرد .

پیاده سازی و رابط کاربری گروه سیستم جمع سپاری ما دارای رابط اختصاصی کاربری برای گروه و کاربر پسند است . رابط کاربری سیستم روی پلت فرم موتور اجرای برنامه گوگل اجرا می شود . آنالیز پشتیبانی پاسخ های گروه و محاسبه پرسش های بعدی به جمعیت ارائه شده در MATLAB R2014b انجام شد. تصویر صفحه رابط کاربری در شکل 4 نمایش داده شده است . ابتداءً ما 20 تصویر را در سمت چپ صفحه نمایش داده ایم (پرسش) و از عضو گروه خواسته شده تصاویر را در 4 صندوق سمت راست کشیده و رها کنند (دراگ اند دراپ) اعضای گروه می توانند تصاویر را اگر به تصاویر دیگر مربوط نباشد بیرون از صندوق قرار دهند در هر مورد الگوریتم ما تنها نشان می دهد که تصاویر باقی مانده دارای کمترین شباهت به تصاویر داخل صندوق هستند. از این رابط کاربری در آزمایشاتی شرح داده شده در ذیل استفاده شده است .

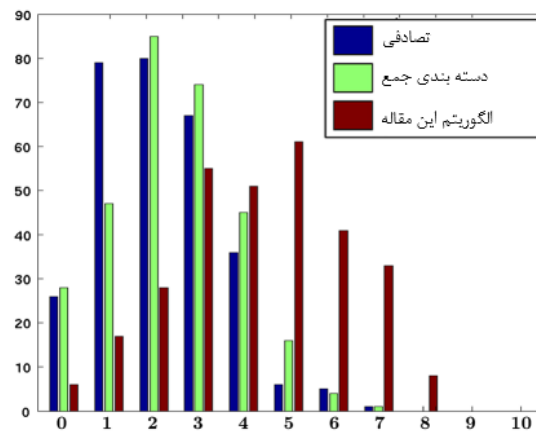


شکل 4. مثالی از رابط کاربری دسته بندی ، a : به کاربر 20 تصویر ارائه شده تا در 4 صندوق سمت راست دسته بندی کند b : صندوق ها ممکن است حاوی تصاویر به حد لازم باشند . وقتی تمام تصاویر جمع بندی شدند ، کاربر پرسشنامه ای تکمیل کرده و 20 تصویر بعدی را دریافت می کند.

#### 4.1 آزمایشات گروه با زمینه میدانی

به عنوان بررسی درستی ما دو آزمایش با مقیاس کوچک با افراد کم (حدود 10 تا 15 عضو گروه) و مجموعه های کوچکی از تصاویر با زمینه میدانی معلوم انجام دادیم . آزمایش ما با دوکار متفاوت محاسباتی صورت گرفت : top-k و دسته بندی . برای هرکار ، اعضای گروه به پرسش های الگوریتم مبنا و الگوریتم ما پاسخ می دهند . رنگهای شبیه top-k ساده ترین مجموعه تصاویری که استفاده کردیم مجموعه ای از 300 رنگ یکدست بود که شباهت آنها با زمینه میدانی قابل اندازه گیری است مثلاً با قرار دادن رنگ ها در فضای سه بعدی برحسب مقدار RGB یا HSL (ما از RGB استفاده خواهیم کرد) هدف محاسبه k-NN شبیه ترین رنگها برای تغییر دادن مقادیر k می باشد. ما نتایج الگوریتم خودمان را با نتایج الگوریتم تصادفی مبنا و الگوریتم های دسته بندی گروه با استفاده از همان تعداد پرسش در تمام الگوریتم ها مقایسه کردیم .

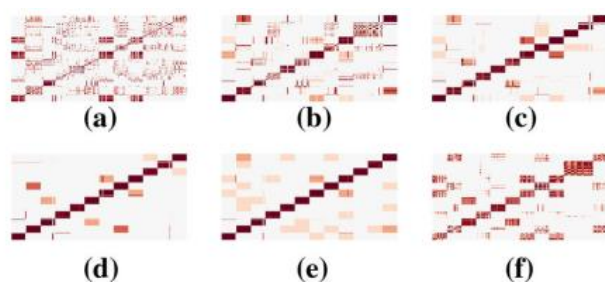
نتایج حاکی از آن است که الگوریتم ما درصد بالاتری از نزدیکترین همسایه ها برای درصد بالاتر تصاویر شناسائی می کند. شکل 5 نشان دهنده ده نتیجه NN در سه الگوریتم با استفاده از 235 سؤال در کل است. پنج مرحله در الگوریتم ما استفاده شد. برای هر الگوریتم ما نمودار ستونی مقطع بین 10-NN (بر طبق زمینه میدانی) و 10-NN محاسبه شده را نمایش دادیم. توجه نمائید که دسته بندی گروه اندکی بهتر از الگوریتم مبنای تصادفی است، اما الگوریتم ما در کل بخش وسیعی از تصاویر 10-NN را شناسائی می کند بالاتر ستون های سمت راست نمودار (ستون های قرمز). در کل الگوریتم ما 50-43.4٪ بیشتر از الگوریتم های مبنای دیگر نزدیکترین همسایه های واقعی را شناسائی می کند که حاکی از کارآمدی روش بهبود مستمر ما می باشد.



شکل 5 – تعداد 10-NN صحیح براساس داده گروه واقعی، مقایسه نتایج الگوریتم ما با دو گزینه مبنا

**دسته بندی فونت ها** در این آزمایش ما توانائی الگوریتم خود را برای دسته بندی نامه تصاویر برحسب فونت درحالتی که زمینه میدانی فونت متعلق به نامه ها بود، آزمودیم. ما از 180 نامه با 12 نوع فونت مختلف استفاده کردیم و از اعضای گروه خواستیم تا شباهتهای نامه ها از جنبه ظاهری آنها ارزیابی کنند. نتایج برای محاسبه 12 دسته نامه که به صورت مطلوب و دقیق با 12 نوع فونت اصلی تطبیق داشت مورد استفاده قرار گرفت. الگوریتم ما در کل 5 مرحله اصلاح از 123 سؤال استفاده نمود. برای مقایسه ما همان کار را با 123 سؤال تصادفی انجام دادیم. نتایج الگوریتم تقریباً عالی بود 1.1% خطا (دو نامه). در مقایسه انتخاب تصادفی سؤال دارای خطای تقریبی 60٪ بود، و همچنین توسط الگوریتم ما بعد از مرحله دوم عملکرد بهتری حاصل شد. شکل 7 خوشه نمونه ایجاد شده با الگوریتم ما و خوشه مرتبط ایجاد شده توسط الگوریتم تصادفی مبنا را نمایش می دهد. خوشه نامه در صورتی که فقط

شامل فونت های دست نویس باشد زمینه وسیع تری از فونت ها را حس می کند ، ؛ اما بهبود مستمر در روش ما امکان تشخیص فونتهای مختلف دست نویس را ایجاد می کند .



شکل 6 . نقشه های حرارتی نشان دهنده دقت دسته بندی برای مجموعه داده فونت **a** تا **e** حاکی از کیفیت خوشه از فاز 1 تا 5 الگوریتم به ترتیب و در کل 123 سؤال می باشد. برای مقایسه شکل **f** نشان دهنده کیفیت خوشه بعد از 123 سؤال تصادفی است .

<i>a b d f g k m n</i>	<i>r t y r a d f</i>
<i>p q r s t u y</i>	<i>g k q s t b d</i>
<b>(a)</b>	<b>(b)</b>

شکل 7 . دو مثال از خوشه های ایجاد شده برای یک نامه “a” (بالا سمت چپ) بر اساس معیار شباهت **a** الگوریتم ما و **b** الگوریتم تصادفی مینا

## 4.2 آزمایشات گروه با مجموعه داده های دنیای واقعی

در مرحله بعدی ما آزمایشاتی با دو مجموعه داده دنیای واقعی ترتیب دادیم و این در حالی انجام شد که شباهت تصاویر به مقدار بسیار زیادی معنایی و به تبع آن حالتهای تصویر برای برآورد این شباهت ناکافی بود . اولین مجموعه داده شامل 910 تصویر از پوستره های فیلم بود که از صفحات فیلم از ویکی پدیا دانلود شده بود ، شباهت بر اساس ژانر ، استایل پوستر ، کاراکترها و غیره بود . برای این مجموعه ما از حدود 60 عضو گروه 547 سؤال جمع آوری کردیم .

مجموعه داده دوم شامل 1024 صندلی از انواع و زاویه های متفاوت از مجموعه داده ShapeNet [3] بود. شباهت ها در این مجموعه داده از میان سایر شباهتها براساس حالت های معنایی است مانند کاربرد صندلی ها ، موادی که احتمالاً از آن ساخته شده و سطح بررسی شده راحتی . برای این مجموعه ما 559 جواب سؤال از حدود 60 نفر از اعضای گروه جمع آوری کردیم .

در برنامه زندگی واقعی برای این مجموعه ها زمینه میدانی یا واحد طلا وجود نداشت . از این رو ما به صورت دستی نتایج را با نمونه گیری تصاویر با k-NN تصاویر آزمودیم و این نتایج را با نتایج وسائل اتوماتیک بر اساس حالت های تصویر مقایسه نمودیم . در مجموعه داده فیلم از نمودار ستونی رنگ با 64 صندوق استفاده کردیم (چهار صندوق برای برای هر کانال RGB) و یک تامبنیل چهار در چهار پیکسل یا در مجموع مقادیر 16 RGB. دو توصیفگر پی در پی بودند و برای محاسبه فاصله به عنوان یک بردار عمل نمودند. برای مجموعه داده صندلی ما از حالت های اخذ شده از توصیفگر HoG بهره بردیم [4].

برای بررسی دستی ، ما از 50 تصویر دانه از هر مجموعه داده استفاده کردیم . برای هر تصویر دانه ما تصاویر 10 NN آن را طبق الگوریتم ما و الگوریتم مبتنی بر ویژگی مبنا نمونه گیری کردیم . به هر کدام از تصاویر با توجه به تصویردانه همان تصویر برچسب " خیلی شبیه " ، " شبیه " و " نامرتبط " زده شد. ما درصد تصاویر دانه که الگوریتم ما تعداد بیشتری تصویر مشابه نسبت به الگوریتم مبنا یافته بود با شکستن گره از تعداد تصاویر بسیار مشابه حساب کردیم . نتایج در ستون درصد موفقیت (Success%) جدول 1 نمایش داده شده است . برای مقاردهی اینکه چقدر ما از الگوریتم مبنا بهتر عمل کرده ایم میانگین تفاوت بین تعداد تصاویر مشابه را نیز محاسبه کردیم . این تفاوت در ستون  $\Delta$  مشخص شده است .

**Table 1** Real-world dataset results

Dataset	Number of images	Success (%)	$\Delta$
Movie posters	910	87.2	2.5
Chairs	1024	76.2	3

ما با استفاده از مثال مشخصی تفاوت های مشاهده شده در شکل 1 را شرح خواهیم داد. شکل نشان دهنده تصاویر-10 NN است (a) که با الگوریتم ما براساس پرسش های دسته بندی و (b) مطابق با تصیفگرهای تصویر است . در این

مورد نتایج برآورد شباهت معنایی ما فیلم هائی با همان ژانر را بازیابی کرده است (فیلم های کارتونی ماجرائی). در داخل آن ژانر اغلب نزدیکترین همسایه ها (4 تا از 5 تصویر بالائی) دارای ظاهر بصری مشابه تصویر دانه هستند (زمینه آبی). از طرف دیگر فیلم های بازیابی شده با استفاده از توصیفگرهای تصویر دارای ظاهر بصری مشابه از لحاظ طرح رنگی و حالت هستند اما از لحاظ معنایی بسیار متفاوت می باشند. توجه کنید هنگامی که از توصیفگرهای تصویر ساده تر و یا حتی از توصیفگرهای پیچیده استفاده می کنیم نمی تواند پوسترهای فیلم مربوطه را در ژانر تصویردانه که دارای نمود بصری یکسان هستند .



شکل 8. نزدیکترین همسایه های تصویر مرکزی در مجموعه صندلی ها (a) با استفاده از توصیفگر HoG و (b) با استفاده از پرسش های گروه محاسبه گردید . تصاویر کوچکتر نشان دهنده همسایه های دورتر است . تصاویر دارای شباهت کمتر های لایت شده اند .

شکل 8 نتایج مشابه برای مجموعه داده صندلی را براساس نتایج k-NN (a) که با استفاده از توصیفگر HoG محاسبه شده اند را نشان می دهد . تصویر دانه صندلی مدرسه با لوله های خم شده متصل به پشت است. صندلی های 10-NN حاصل شده با الگوریتم ما تماماً صندلی مدرسه بوده و بسیاری از آنها دارای اجزای استایل مشابه نظیر لوله های خم شده هستند . در مقابل صندلی های محاسبه شده با توصیفگر HoG به نظر دارای ظاهر مشابه هستند (و نیز دارای ترتیب یکسان هستند) حتی شامل صندلی های دفتری و صندلی های اتاق نهار خوری می باشند و استایل آنها بسیار متفاوت است (عکس های دارای کمترین شباهت در شکل های لایت شده اند ) .

شکل 9 . چند انتخاب از نتایج k-NN در مورد صندلی ها و پوسترهای فیلم نمایش می دهد . در هر مجموعه تصویر بالا سمت چپ تصویر دانه است و 7 همسایه نزدیک آن از چپ به راست قرار داده شده اند. در بسیاری موارد شباهت بین تصاویر می تواند معنایی یا بصری باشد . با موارد را با دقت انتخاب کردیم که رابطه معنایی محض



وجود داشته باشد تا ضبط آن توسط توصیفگرهای تصویر بسیار سخت یا غیر ممکن باشد. پیوند معنایی بین پوسترهای فیلم بسیار متغیر است و شامل فیلم هائی از همان ژانر (a) پوسترهایی که عناصر عمده چاپی دارند (b) پوستر فیلم های قدیمی (c) یا همان قیافه ها در یک پوستر (d) می باشد. پیوند معنایی بین صندلی ها ممکن است اجزای استایل مشابه (e) شکل کلی مشابه (g) یا حتی چرخدار باشد (h). نتایج k-NN برای تمام پوسترهای فیلم و صندلی ها را می توان در مجموعه داده در تکمیلی قابل مشاهده است.



شکل 9. نمونه k-NN تصاویر از مجموعه داده های پوستر فیلم و صندلی

### 4.3 آزمایشات ساختگی

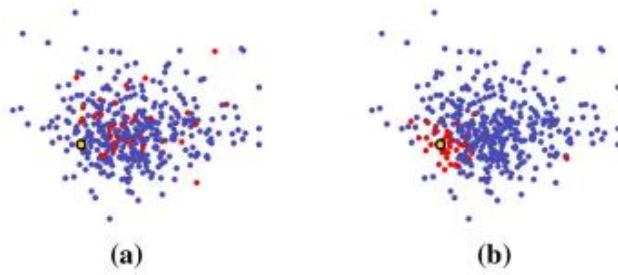
ما بعداً تحلیل از الگوریتم خود را از طریق نتایج آزمایشات ساختگی ارائه خواهیم داد. آزمایشات روی مجموعه داده هائی با زمینه میدانی موجود و نتایج از گروه شبیه سازی شده ترتیب داده شد. پاسخ شبیه سازی شده برای پرسش ارائه شده با استفاده از الگوریتم k-means که 20 تصویر را برای پرسش در 4 دسته جدا می سازد. با استفاده از پاسخ خای ساختگی به ما امکان می دهد عملکرد الگوریتم را در انواعی از حالتها تست کنیم.

**اثر محل** در مقدمه ما به اهمیت استفاده از پرسش ها درباره همسایگی های محلی تصاویر تاکید کردیم. برای تست جداگانه این ادعا ما یک آزمایش ساختگی اختصاصی ترتیب دادیم که به ترتیب زیر است. از 1000 رنگ که

به صورت همسان نمونه گیری شده بودند استفاده کردیم . در حالی که شباهتهای واقعی برای این مجموعه عکس معلوم بود توانستیم محل پرسش ها را تغییر دهیم : برای هر پرسش از تصویر دانه آغاز کردیم سپس باقی تصاویر را از فاصله مشخصی از تصویر دانه نمونه برداری کردیم . سپس از نتایج پرسش ها برای محاسبه جادهی به طور معمول بهره بردیم . ما کاهش خطی در میانگین دقت تصاویر 10-NN را همگام با افزایش فاصله بین تصاویر در هر نمونه مشاهده نمودیم .

**رخداد تصاویر مشابه** یکی از نشانه های کارآمدی بهبود مستمر در الگوریتم ما رخداد تصاویر مشابه در یک پرسش است . در حالت مطلوب معیار شباهتی که ما محاسبه کردیم هر قدر همگرا باشد به همان میزان احتمال حضور تصاویر مشابه در یک پرسش در کنار هم بیشتر است . علاوه بر این انتظار این است که فاصله بین دوگانه هائی که در بسیاری از پرسش ها با هم ظاهر می شوند ، در صورت دسترسی به داده بیشتر ، دقیق باشد . در حالی که بودجه پرسش ها محدود است ، تکرار هر دوگانه پرسش شده همراه با هزینه مجدد خواهد بود و نتیجه آن کمتر بودن اطلاعات موجود است . ما نشان خواهیم داد که الگوریتم ما توجه موثری بر دوگانه هائی دارد که نزدیک به هم هستند و بنابر این نیاز به اطلاعات دقیق تری داریم .

شکل 10 حاکی از این مسئله است . ما تصاویر دو بعدی تعبیه شده را شبیه سازی کردیم و این به صورتی است که هر نقطه نماینده یک تصویر در مجموعه داده است . فاصله بین دوگانه نقاط (یا تصاویر) از جادهی اخذ شده که شباهت زمینه میدانی را شبیه سازی می کند. مجموعه داده حاوی 400 عکس است و ما 400 پرسش شبیه سازی شده ارائه دادیم ، یک بار از الگوریتم ما و یک بار با استفاده از پرسش های تصادفی استفاده شد. سپس یک تصویر دلخواه علامتگذاری کردیم (طلائی) و تعداد دفعاتی که هر تصویر به همراه آن حضور یافت را شمردیم . ما تصاویر را بر حسب شمارش پرسش های دوسویه امتیاز بندی کردیم . 20 تصویر دارای بالاترین امتیاز (5٪ مجموعه داده) که با هم مورد پرسش قرار گرفته بودند را به رنگ قرمز روشن در آوردیم . 20 تصویر بعدی (5٪) به رنگ قرمز تیره درآمدند . باقی تصاویر (36 یا 90٪) به رنگ آبی روشن درآورده شدند.



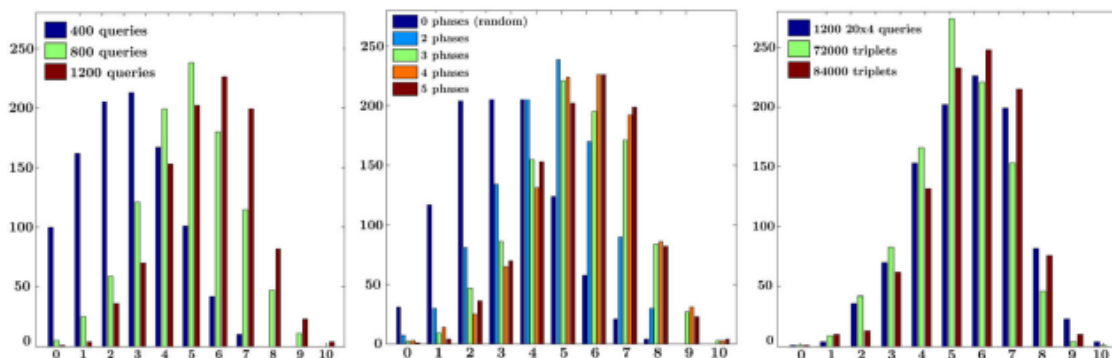
شکل 10. نمایش ظاهری تصاویر که در یک پرسش حضور داشته اند به صورت تصویر علامتگذاری شده به رنگ پلائی . تصاویر بر حسب تعداد پرسش های دوسویه امتیاز دهی کردیم و 10٪ تصاویر دارای امتیاز بالا را با رنگ قرمز نشان دادیم a پرسش های دوسویه بعد از 400 پرسش تصادفی b پرسش های دوسویه بعد از 400 پرسش با استفاده از الگوریتم ما

شکل 10 نشان می دهد که با استفاده از انتخاب تصادفی پرسش ، تصاویری که با تصویر دارای علامت پلائی همزمان ظاهر شده اند ، طبق انتظار به صورت تصادفی پخش شده اند . در مقابل با استفاده از الگوریتم ما در انتخاب پرسش ها (شکل 10 b) ، تصاویری که به تکرار باهم ظاهر شده اند در حوالی تصویری پلائی ظاهر شده اند . آشکار است که ما برای پرسش هایی که معلوم است دور از هم هستند هزینه نکرده ایم ، در حالی که فاصله آنها از همدیگر اهمیت کمتری دارد و انتظار می رود دقت کمتری داشته باشند . این امر به الگوریتم ما این امکان را می دهد تا شباهتهای مکانی را بهتر برآورد نماید . و از آنها برای تخمین شباهتهای کلی استفاده کند .

**تغییر دادن پارامترهای الگوریتم** در بخش بعدی الگوریتم خود را با تغییر مقدار دو پارامتر اجرا خواهیم کرد : تعداد کل پرسش ها و تعداد مراحل برای به نمایش گذاشتن تاثیر این دو بر نتایج پرسش ها. شکل 11 (راست) نشان دهنده تاثیر تغییر دادن تعداد کل پرسش ها در می باشد ، در مجموعه داده ساختگی 1000 رنگ تصادفی و 5 مرحله از الگوریتم ما . مطابق انتظار بین تعداد پرسش ها ئی که مورد استفاده قرار گرفت و کیفیت نتایج ، ارتباط مثبتی وجود دارد و این ارتباط با اندازه انطباق 10-NN واقعی تصاویر و 10-NN محاسبه شده توسط ما اندازه گیری می شود . این بدین معناست که اگر بودجه بیشتری در اختیار داشتیم می توانستیم برآورد معیار شباهت را بهبود بخشیم . شکل 11 (مرکز) حاکی از تاثیر مراحل روی کیفیت نتایج است ( با استفاده از مجموعه تصویر به طریقی که در بالا

ذکر شد ، معیار شباهت یکسان و در کل 1200 پرسش). تعداد از صفر (که برابر با انتخاب تصادفی پرسش است ) تا 5 تغییر می کند . توجه کنید افزایش تعداد مراحل کیفیت نتیجه را بالا می برد، همراه با این که محاسبه مجدد جادهی اغلب امکان ایجاد پرسش های بهتر را ایجاد می کند. با این حال اندازه ای که نتایج بهتر می شود با 4 مرحله بودن و یا 5 مرحله ای بودن بسیار کم است .

پرسش ها بر روی تصاویر سه گانه راه حل معمول برای جمع آوری مقایسات تصاویر از گروه بر اساس سئوالات سه گانه با فرم " آیا تصویر A به تصویر B شبیه است یا C " می باشد. ما قبلاً به این نکته توجه داشته ایم که مزیت روش ما روی تصاویر سه گانه آن است که دسته بندی سئوالات زمینه را برای مقایسه فراهم می کند . در این آزمایش ساختگی از زمینه چشم پوشی کردیم و روی تعداد سئوالات لازم بر هر نوع از راه حل تمرکز یافتیم . به گونه ای که در شکل 11 (راست) نشان داده شده ، عملکرد الگوریتم ما با استفاده از 1200 پرسش قابل مقایسه با الگوریتم ها براساس پرسش سه گانه است که از 84000 پرسش استفاده نموده اند .



شکل 11. تعداد NN-10 صحیح تصاویر به صورت تابعی از پرسش ها (چپ) و تعداد مراحل (مرکز) در مقایسه با الگوریتم

های مبتنی بر پرسش سه گانه

## 5. نتیجه گیری

در این مقاله ، ما روش موثری برای برآورد شباهتهای تصاویر منحصراً بر مبنای داده های گروه ارائه دادیم . سیستم ما به صورت مستمر تصاویر ارائه شده به گروه را تصحیح می کند تا مقایسات شباهت بین تصاویر در یک همسایگی

را بدست آورد که همگرایی برای معیار دقیق شباهت را ممکن سازد. در بررسی آزمایشی ما تعداد به ویژه کمی از پرسش ها مورد استفاده قرار گرفت بر این اساس که بتوانیم برآورد نسبتاً خوبی از شباهت معنایی حاصل نمائیم .

**محدودیت ها و کار بعدی** این کارتنها بر داده های ارائه شده توسط گروه تمرکز یافته بود . با این حال به

تکرار در این مورد ذکر شد ، سررشته ای که برای شباهت معنایی تصاویر وجود دارد در فرم حالت های تصاویر یا بافت متنی موجود است . حتی اگر این سرخخا برای کل محدوده ارتباطات معنایی در حساب نیاید ، جالب خواهد بود که

چگونه آنها را در پیوند همزمان با الگوریتم ما مورد استفاده قرار دهیم. این جنبه در مقیاس پذیری روش ها مخصوصاً

در مجموعه های بسیار بزرگ تصویر ، تعداد مقرون به صرفه پرسش ها حتی ممکن است در اندازه مجموعه خطی

نباشد. یک روش درست برای تجمیع سررشته های معنایی، از الگوریتم ما برای آموختن شباهتها برای بخش کوچکی

از مجموعه تصویر استفاده خواهد نمود و سپس با به کارگیری تکنیک های آموزش ماشین برای تکمیل بقیه کار ، از

حالت های سررشته های معنایی استفاده خواهد نمود. (ایجاد شده توسط Lun و دیگران [9] ، Saleh و دیگران

[13]، Yi و دیگران [22]). راه حل جالب تر ممکن است ترکیب سررشته ها با مرحله ایجاد پرسش باشد . چون

بهره بردن از سایر برآوردها می تواند به صورت بالقوه باعث نادیده شدن تصاویر دارای شباهت معنایی شود ، این امر

کم اهمیت نیست. روش چالش برانگیز دیگر برای کارآینده طرز برخورد دقیق تر با عدم قطعیت ناشی از عمل گروه باشد.

اغلب اعضای گروه روی شباهت تصاویر هم رای نیستند، یا حتی جواب های متناقضی ارائه می دهند. تا این مرحله ما

فرض کردیم که جادهی که انجام دادیم اثر چنین تناقض هائی را کاهش می دهد . با این حال ، شاید بخواهیم برای

تناقض ها با مدلینگ احتمال گرایانه رفتار گروه محاسبه دقیقی داشته باشیم مثلاً به گونه ای که در [7] به منظور

دسته بندی انجام شده است . بنابر این جالب خواهد بود که مدل های احتمال گرایانه را اختصاصاً برای یادگیری

معیار شباهت توسعه دهیم . بویژه ، این روش باید محاسبات کارآمد را هم کنش با ماهیت هم کنشی الگوریتم ما

پشتیبانی کند.

## References

1. Bar-Hillel, A., Hertz, T., Shental, N., Weinshall, D.: Learning a mahalanobis metric from equivalence constraints. *J. Mach. Learn. Res.* 6(6), 937–965 (2005)
2. Biswas, A., Jacobs, D.: Active image clustering with pairwise constraints from humans. *Int. J. Comput. Vis.* 108(1–2), 133–147 (2014)
3. Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., et al.: Shapenet: An information-rich 3d model repository. arXiv:1512.03012 (arXiv preprint) (2015)
4. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: *Computer vision and pattern recognition, IEEE, 2005*, vol. 1, pp. 886–893 (2005)
5. Davidson, S.B., Khanna, S., Milo, T., Roy, S.: Using the crowd for top-k and group-by queries. In: *International conference on database theory*, pp. 225–236. ACM (2013)
6. Frome, A., Singer, Y., Sha, F., Malik, J.: Learning globallyconsistent local distance functions for shape-based image retrieval and classification. In: *International conference on computer vision, IEEE*. pp. 1–8 (2007)
7. Gomes, R.G., Welinder, P., Krause, A., Perona, P.: Crowdclustering. In: *Advances in neural information processing systems*. pp. 558–566 (2011)
8. Lowe, D.G.: Object recognition from local scale-invariant features. In: *International conference on computer vision, IEEE 1999*, vol. 2, pp. 1150–1157 (1999)
9. Lun, Z., Kalogerakis, E., Sheffer, A.: Elements of style: learning perceptual shape style similarity. *ACM Trans. Gr. (TOG)* 34(4), 84 (2015)
10. Marcus, A., Wu, E., Karger, D., Madden, S., Miller, R.: Humanpowered sorts and joins. *Proc. VLDB Endow.* 5(1), 13–24 (2011)
11. O’Donovan, P., L̄ibeks, J., Agarwala, A., Hertzmann, A.: Exploratory font selection using crowdsourced attributes. *ACM Trans. Gr. (TOG)* 33(4), 92 (2014)
12. Oliva, A., Torralba, A.: Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int. J. Comput. Vis.* 42(3), 145–175 (2001)
13. Saleh, B., Dontcheva, M., Hertzmann, A., Liu, Z.: Learning style similarity for searching infographics. In: *Proceedings of the 41st Graphics Interface Conference*, pp. 59–64. Canadian Information Processing Society (2015)
14. Sammon, J.W.: A nonlinear mapping for data structure analysis. In: *IEEE transactions on computers* (1969)
15. Sivic, J., Zisserman, A.: Video google: a text retrieval approach to object matching in videos. In: *International conference on computer vision, IEEE 2003*, pp. 1470–1477 (2003)
16. Tamuz, O., Liu, C., Shamir, O., Kalai, A., Belongie, S.J.: Adaptively learning the crowd kernel. In: *International conference on machine learning (ICML-11)*, pp. 673–680. ACM (2011)
17. Wang, C., Blei, D., Li, F.-F.: Simultaneous image classification and annotation. *Computer vision and pattern recognition, IEEE 2009*, pp. 1903–1910 (2009)
18. Wang, J., Kraska, T., Franklin, M.J., Feng, J.: Crowder: crowdsourcing entity resolution. *Proc. VLDB Endow.* 5(11), 1483–1494 (2012)
19. Weinberger, K.Q., Blitzer, J., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. In: *Advances in neural information processing systems*, pp. 1473–1480 (2005)
20. Wilber, M.J., Kwak, I.S., Belongie, S.J.: Cost-effective hits for relative similarity comparisons. In: *Conference on human computation and crowdsourcing* (2014)
21. Xing, E.P., Ng, A.Y., Jordan, M.I., Russell, S.: Distance metric learning with application to clustering with side-information. *Adv. Neural Inf. Proc. Syst.* 15, 505–512 (2003)
22. Yi, J., Jin, R., Jain, S., Yang, T., Jain, A.K.: Semi-crowdsourced clustering: Generalizing crowd labeling by robust distance metric learning. In: *Advances in neural information processing systems*, pp. 1772–1780 (2012)
23. Zha, Z.-J., Hua, X.-S., Mei, T., Wang, J., Qi, G.-J., Wang, Z.: Joint multi-label multi-instance learning for image classification. *Computer vision and pattern recognition, IEEE 2008*, pp. 1–8 (2008)