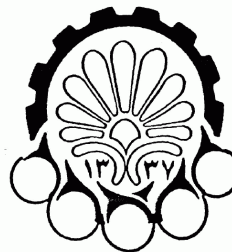




بسمه تعالی



پروژه پایانی درس سیستم های چند رسانه ای

عنوان پروژه:

پردازش معنایی سیگنال های چند رسانه ای

نام استاد: دکتر احدی

ارائه دهنده: سپند حقیقی

شماره دانشجویی: ۹۰۲۳۰۷۶

زمستان ۹۳

فهرست مطالب

چکیده	۵
۱-مقدمه	۶
۲-مفاهیم اولیه و مشکلات پردازش معنایی	۷
۲-۱ شکاف معنایی	۸
۲-شکاف حسگری	۹
۳-دو سوال بنیادی	۱۰
۴- کاربردهای پردازش معنایی	۱۱
۴-۱ پردازش معنایی صوت	۱۱
۴-۲ پردازش معنایی متن	۱۱
۴-۳ پردازش معنایی تصویر	۱۱
۵- پردازش معنایی فیلم	۱۲
۵-۱ استخراج فریم	۱۳
۵-۲ استخراج ویژگی	۱۴
۵-۳ خوشه بندی	۱۸
۵-۴ مدل بردارها	۲۱

۵-۵ آنالیز مفهوم پنهانی ۲۲

۵-۶ طبقه بندی ۲۴

۵-۷ استفاده از روش های فراابتکاری ۲۵

۶- نتایج عملی ۲۶

۷- منابع ۲۸

فهرست اشکال و جداول

شکل-۱	۷	پردازش-آنالیز - درک
شکل-۲	۹	شکاف حسگری
شکل-۳	۹	شکاف معنایی
شکل-۴	۱۲	دیاگرام بلوکی پردازش معنایی فیلم
شکل-۵	۱۳	استخراج فریم های کلیدی
شکل-۶	۱۷	چند فریم متوالی
شکل-۷	۱۷	جداسازی ناحیه شکل-۶ با استفاده از RSST
شکل-۸	۱۹	مثالی از پیاده سازی K-Means
شکل-۹	۲۰	تصویر ابتدایی ماشین
شکل-۱۰	۲۰	نواحی استخراج شده
شکل-۱۱	۲۴	طبقه بندی غیر خطی
جدول-۱	۲۶	نتایج عملی بر روی ۹ ناحیه
شکل-۱۲	۲۷	نتایج تشخیص موضوعی
جدول-۲	۲۷	بازده و دقت الگوریتم

چکیده:

پردازش سیگنال های چند رسانه ای یکی از مباحث مطرح امروزه می باشد که دارای انواع گوناگونی است و پردازش معنایی یکی از پرکاربردترین آنهاست که توانایی تبدیل ویژگی های سطح پایین به ویژگی های سطح بالا و فضای مفهوم را دارد برای مثالی از کاربردهای تجاری آن می توان به جستجو بر اساس عکس در گوگل اشاره کرد.

واژگان کلیدی: پردازش معنایی - معنا - سیگنال - چند رسانه ای - استخراج مفهوم

۱- مقدمه :

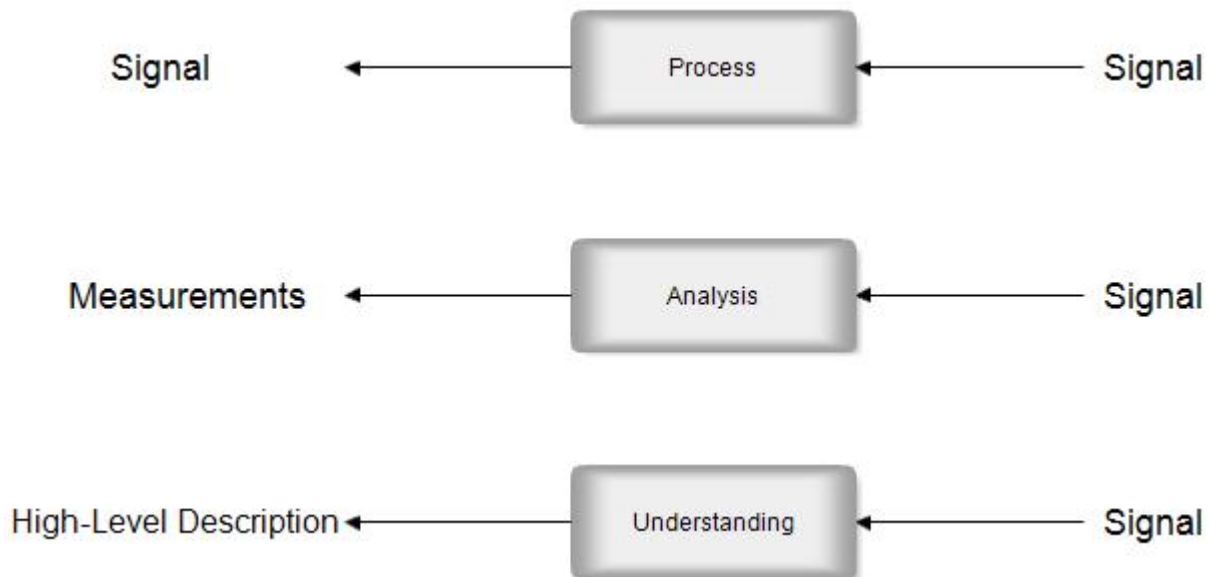
یکی از پرکاربردترین علوم روز دنیا پردازش سیگنال های چند رسانه ای است که از بین آنها می توان پردازش تصویر و پردازش را مثال زد ، در سال های اخیر ، دانشمندان تمام تلاش خود را برای نزدیک تر ساختن این پردازش ها به مفاهیم سطح بالا کرده اند که این حلقه گم شده همان پردازش معنایی است .

پردازش معنایی سیگنال یا معادل لاتین آن **Signal Semantic Processing** در لغت به معنای استخراج معانی سطح بالا یا قابل فهم مستقیم برای انسان از سیگنال است که با توجه به اینکه تمام اندازه گیری ها در علوم پردازش سیگنال مفاهیم عددی و غیر ملموس است تبدیل این مفاهیم به معانی سطح بالا چندان آسان به نظر نمی رسد.

در سال های اخیر الگوریتم های بسیاری برای تبدیل معانی سطح پایین به فضای مفهوم و معانی سطح بالا مطرح شده است . در این مقاله ابتدا مفاهیم ابتدایی پردازش معنایی و کاربرد های آن آورده شده سپس یک الگوریتم برای پردازش معنایی تصویر های ویدیویی که بر مبنای [1],[2],[3] است مورد بررسی قرار گرفته است.

۲- مفاهیم اولیه و مشکلات پردازش معنایی :

به طور کلی ، می توان سه مفهوم پردازش ، آنالیز و درک را به مانند شکل ۱- مورد مطالعه قرار داد :



شکل-۱ (پردازش - آنالیز - درک)

همان طور که در مقدمه اشاره شد ، پردازش معنایی مرحله آخر این شکل یا همان استخراج مفاهیم سطح بالاست ، که با ورودی سیگنال به سیستم خروجی را از فضای مفهوم و توصیفات سطح بالا دریافت می کنیم. در این تبدیل دو مشکل بزرگ به نام شکاف حسگری و شکاف معنایی وجود دارد.[1]

۱-۲ شکاف معنایی :

شکاف معنایی ، به مساله عدم تطبیق میان مفاهیم استخراج شده توسط یک سیستم تحلیل معنایی چند رسانه ای و مفاهیم قابل درک توسط انسان می پردازد . به عبارت دیگر ، تفسیر معانی از اطلاعات چندرسانه ای در سیستم کامپیوتری و انسان متفاوت است ، که این موضوع طراحی سیستم های تحلیل معنایی سیگنال های چندرسانه ای را مشکل می کند.

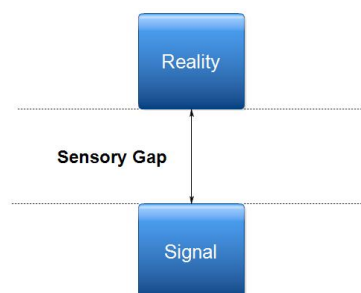
به عنوان یک مثال ساده ، انسان به راحتی می تواند وقوع گل در فوتبال را تشخیص دهد ، اما آشکارسازی گل توسط کامپیوتر نیازمند پردازش های زیادی است . به طور خلاصه این پردازش ها عبارتند از آشکارسازی توپ ، آشکارسازی دروازه ها ، ردیابی توپ و تعیین این که آیا توپ از خط دروازه ها عبور کرد یا خیر . بررسی محتوای یک سیگنال چندرسانه ای و استخراج اطلاعات معنایی از آن مشکلاتی را به دنبال دارد که برای حل آن نیازمند ارائه الگوریتم های کارآمد هستیم . این مشکلات با عنوان شکاف معنایی مطرح می شوند.

با توجه به آنچه گفته شد ، برای استخراج رویداد ها و مفاهیم یک سیگنال ، لازم است پردازش اطلاعات در سطوح معنایی مختلف انجام گیرد. ماهیت پردازش ها در هر سطح متفاوت بوده و در نتیجه اطلاعات استخراج شده از آنها در سطح معنایی متفاوتی خواهد بود. بر اساس یک دسته بندی کلی ، پردازش های انجام شده برای استخراج اطلاعات معنایی از سیگنال را می توان در سه سطح طبقه بندی کرد : ۱- پردازش سطح پایین ۲-پردازش سطح میانی ۳-پردازش سطح بالا برای حذف شکاف معنایی میان ویژگی های سطح پایین و مفاهیم سطح بالا باید سعی نمود استخراج ویژگی ها بر اساس سطوح معرفی شده انجام شود.[1]

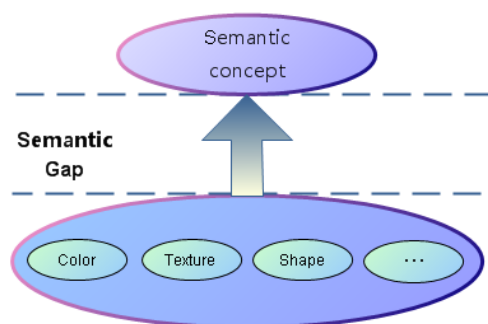
۲-۲ شکاف حسگری :

به شکاف میان موجودیت شی در عالم واقعیت و توصیف آن در سیگنال ذخیره شده ، شکاف حسگری می گویند. به عبارت دیگر شکاف حسگری به مشکلات ابزار های ذخیره سازی و نقش آنها در تحلیل سیگنال می پردازد. برای مثال برای رفع شکاف حسگری در سیستم های تحلیل معنایی ویدیو ، روش ها و ابزار های مختلف تصویر برداری استفاده شده است و ویژگی های مهم در تصویر برداری که ممکن است عدم توجه به آنها باعث ایجاد شکاف حسگری شود عبارتند از :

- ۱- قابلیت کنترل دوربین ، ۲- مکان و زاویه دوربین ، ۳- تعداد دوربین ها ، ۴- سرعت تصویر برداری ، ۵- کیفیت و وضوح تصویر ، ۶- طیف تصویربرداری.[1]



شکل ۲- (شکاف حسگری)



شکل ۳- (شکاف معنایی)[8]

۳- دو سوال بنیادی در پردازش معنایی :

۱- آیا می توانیم چگونگی ایجاد محتویات معنایی را بدست آوریم؟ به عبارت دیگر در جهت معکوس ویژگی های سطح پایین را از روی ویژگی های سطح بالا و مفهوم بازسازی کنیم؟

۲- چگونگی سوال اول را بدون محدودیت موضوعی پاسخ دهیم؟ برای مثال چگونه سیستمی طراحی کنیم که تنها برای یک ورزش خاص کار نکند بلکه برای گروهی بزرگتر از ورزش ها عمل کند؟

این دو سوال هنوز به طور کامل پاسخ داده نشده اند ، ولی یک راه حل ساده آن شناخت ویژگی های ثابت و متغیر در موضوعات مورد بررسی و استفاده درست از آنها در فرآیند پردازش معنایی است.

برای مثال تصور کنید می خواهیم سیستمی طراحی کنیم که برای تشخیص رویداد در کلیه ورزش های میدانی استفاده شود برای این کار باید از ویژگی های مشترک این ورزش ها استفاده شود تا سیستم تشخیص نسبت به نوع ورزش مقاوم تر باشد در زیر نمونه ای از ویژگی های مشترک ورزش های میدانی آمده است :

۱- تشکیل از دو تیم + داور ۲- زمین بازی بسته ۳- زمین بازی چمن+خط کناری ۴- صدای گزارش گر و ... [3]

۴- کاربرد های پردازش معنایی :

۴-۱ پردازش معنایی متن :

پردازش معنایی متن در واقع ابتدایی ترین و نقطه شروع پردازش معنایی است که در علم پردازش زبان های طبیعی (Natural Language Processing) بررسی می شود و کاربرد های مهم آن در گرفتن غلط های نگارشی و املائی در برنامه های ویرایشی متن مثل Microsoft Word و ... و همچنین تجزیه و تحلیل متن برای پاسخ گویی در سیستم های هوشمند مثل LexaLYTICS است.

۴-۲ پردازش معنایی صوت :

پردازش معنایی صوت که یکی از شاخه های علم Speech And Audio Processing محسوب می شود امروزه کاربرد های بسیار زیادی دارد که بیشتر آن ها در سیستم های پاسخگویی هوشمند مانند جستجوی صوتی گوگل یا برنامه هوشمند Siri شرکت Apple است که با استفاده از تجزیه تحلیل صدا و استخراج معنای آن واکنش مناسب را انجام می دهند.

۴-۳ پردازش معنایی تصویر :

پردازش معنایی تصویر و ویدیو یکی از شاخه های Image And Video Processing محسوب می شود . یکی از مثال های بسیار مشهور آن جستجو با استفاده از عکس گوگل است و برای سایر استفاده های آن می توان به کاربرد آن در صنایع نظامی خودکار برای تشخیص مواضع دشمن و تقسیم بندی نواحی تصویر اشاره کرد.

۵- پردازش معنایی فیلم :

در این قسمت ، یک روش برای پردازش معنایی فیلم و مراحل مختلف آن را بررسی می کنیم.[1]
این روش که مبنای آن استفاده از روش آنالیز پنهان مفهومی (LSA) است از ۷ مرحله زیر تشکیل شده است :

۱- استخراج فریم های کلیدی

۲- استخراج ویژگی های سطح پایین

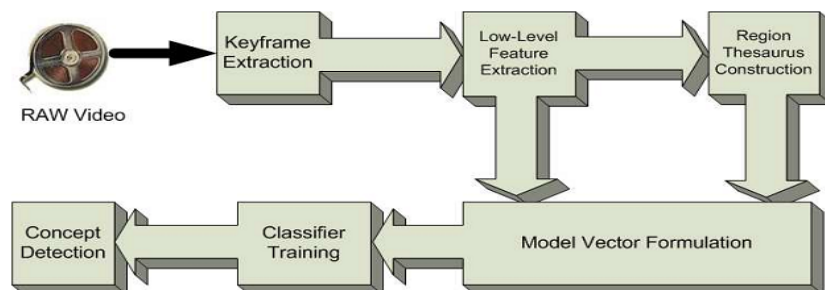
۳- خوشه بندی ناحیه ها

۴- مدل بردارها

۵- آنالیز مفهوم پنهانی (LSA)

۶- استفاده از Classifier

۷- استفاده از روش های فرا ابتکاری [3]



شکل-۴ (دیاگرام بلوکی پردازش معنایی فیلم)[2]

۱-۵ استخراج فریم های کلیدی :

با توجه به اینکه پردازش یک فیلم به صورت پردازش سکانس های مختلف آن صورت می گیرد اولین مرحله پردازش ، استخراج سکانس های آن به صورت مناسب است.

فریم های کلیدی ، فریم های هستند که ابتدا و انتهای یک حرکت و یا تغییر جهت آن را نشان می دهد به عبارت دیگر با استفاده از فریم های کلیدی می توان یک سکانس از یک فیلم را بازسازی کرد یکی از مزیت های این مرحله جلوگیری از پردازش فریم های اضافی است.[6]



شکل-۵ (فریم های کلیدی)[6]

۲-۵ استخراج ویژگی های سطح پایین:

مرحله بعدی استخراج ویژگی های سطح پایین فریم های کلیدی فیلم است ، در این مرحله ابتدا فریم های کلیدی را با استفاده از الگوریتم RSST به نواحی مختلف تقسیم کرده سپس ۶ ویژگی سطح پایین زیر را که ویژگی های بنیادی Mpeg-7 است ، برای هر ناحیه استخراج می کنیم.

۱-رنگ غالب

۲-ساختار رنگ

۳-طرح رنگ

۴-هیستوگرام لبه

۵-همگنی بافت

۶-مقیاس بندی رنگ

$k_i \in K$ نشان دهنده فریم کلیدی شات $s_i \in S$ است ، که S و K مجموعه تمام فریم های کلیدی و شات های فیلم است و در حالی که R مجموعه تمام ناحیه ها است ، $r_i \in R$ نشان دهنده ناحیه ام و $R(k_i) \subset R$ نشان دهنده تمام ناحیه های فریم کلیدی k_i است و در انتها بردار ویژگی یک ناحیه را با $f(r)$ نشان می دهیم.[3]

۱-۲-۵ الگوریتم RSST:

الگوریتم RSST (Recursive Short Spanning Tree) الگوریتمی مبتنی بر گراف است که برای ناحیه بندی تصاویر نیز از آن استفاده می شود از مزیت های این روش می توان به بدون پارامتر بودن ، سریع بودن و عدم نیاز به حدس اولیه اشاره کرد. در این قسمت این الگوریتم را به صورت مختصر تشریح می کنیم :

هدف از ناحیه بندی استخراج نواحی متصل به هم و بدون همپوشانی است ، به صورتی که برای هر ناحیه R_i مجموعه از پارامترها ی $\theta(R_i)$ وجود داشته باشد به طوری که این پارامتر ها بردار حرکت $v(x, y)$ را به صورت کامل تشریح کنند. برای مدل حرکتی دو بعدی که برای ناحیه بندی تصویر نیز مناسب است مجموعه پارامتر ها به صورت $\theta(R_i) = \{S_i, T_i\}$ است که در آن S_i ماتریس دوران با ابعاد 2×2 و T_i بردار ترجمه با ابعاد 2×1 است و بردار حرکت به صورت زیر تقریب زده می شود :

$$w_{R_i}(x, y) = S_i \begin{bmatrix} x \\ y \end{bmatrix} + T_i$$

و برای یک ماسک ناحیه بندی ثابت ، خطای تقریب زدن از رابطه زیر بدست می آید :

$$D = \sum_{i=1}^K D(R_i) = \sum_{i=1}^K \sum_{(x, y) \in R_i} \|v(x, y) - w_{R_i}(x, y)\|^2$$

در الگوریتم RSST یک ماسک ناحیه بندی به صورت یک گراف وزن دار مدل می شود که در آن گره ها به عنوان ناحیه در نظر گرفته می شود و هر جفت از ۴ ناحیه های همسایه (R_i, R_j) با یک یال به نام L_{ij} به هم متصل شده اند.

در ابتدا تصویر به بلوک های 2×2 تقسیم می شود و هر بلوک به عنوان یک ناحیه در نظر گرفته می شود در هر مرحله اگر شرط زیر برقرار باشد یال $L_{i^*j^*}$ از گراف حذف شده و دو ناحیه R_{i^*}, R_{j^*} با هم ادغام می شوند :

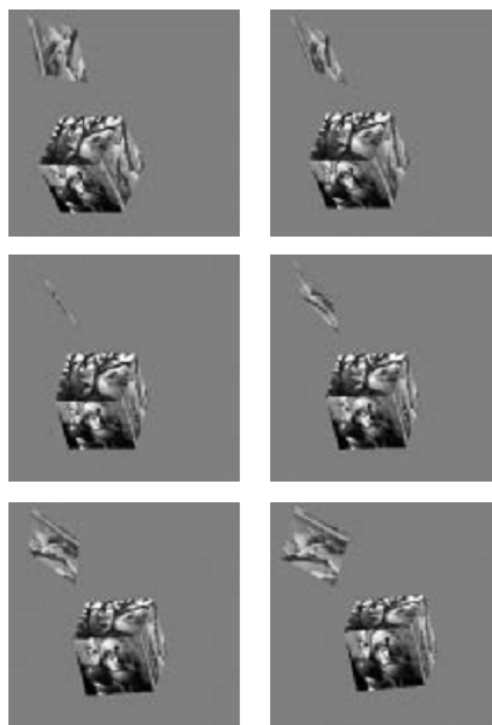
$$(i^*, j^*) = \arg \min_{i, j} d_{ij}$$

و وزن یال های L_{i^*x}, L_{xj^*} که ناحیه ادغام شده $R_{i^*} \cup R_{j^*}$ را به سایر گره ها متصل می کند دوباره محاسبه می شود با تکرار این مراحل به صورت پی در پی تا رسیدن به یک ناحیه واحد ، یک درخت سلسه مراتبی جداسازی ناحیه ها ایجاد می شود که K امین سطح این درخت ، ماسک جداسازی برای K ناحیه است.

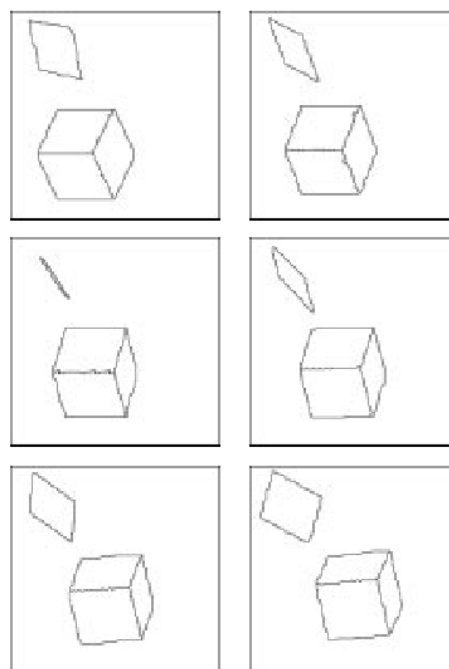
برای حداقل سازی D (خطای تقریب سازی) ، یک راه حل منطقی محاسبه وزن d_{ij} به صورت زیر است :

$$d_{ij} = \sum_{(x,y) \in R_i \cup R_j} \|v(x,y) - w_{R_i \cup R_j(x,y)}\|^2 - \sum_{(x,y) \in R_i} \|v(x,y) - w_{R_i(x,y)}\|^2 - \sum_{(x,y) \in R_j} \|v(x,y) - w_{R_j(x,y)}\|^2$$

که در این عبارت d_{ij} دقیقا برابر ، مقدار افزایش D بر اثر ادغام دو ناحیه R_i, R_j قرار داده شده است . [7]



شکل-۶ (چند فریم متوالی) [7]



شکل-۷ (جداسازی ناحیه شکل ۶- با استفاده از RSST) [7]

۳-۵ خوشه بندی ناحیه ها :

در این بخش با استفاده از الگوریتم های خوشه بندی مثل K-Means اقدام به خوشه بندی نواحی ایجاد شده اولیه بر روی فریم های کلیدی می کنیم. تعداد خوشه ها که با N_T نمایش داده می شود به صورت تجربی به دست می آید. این خوشه بندی بر روی بردار ویژگی ها با استفاده از فاصله اقلیدسی برای معیار اندازه گیری شباهت خوشه های مختلف انجام می شود. نواحی که به مرکز خوشه های حاصل نزدیک تر هستند برای تشکیل دیکشنری نواحی استفاده می شوند ، به این نواحی که با $w_i, i=1, \dots, N_T$ نمایش داده می شوند گونه های نواحی گفته می شود. کاملاً مشخص است که گونه های نواحی ، حاوی اطلاعات مفهومی و معنایی نمی باشند اما در مقایسه با ویژگی های ساختاری و رنگی دارای اطلاعات سطح بالاتری هستند [3].

۱-۳-۵ الگوریتم K-Means :

این الگوریتم برای خوشه بندی استفاده می شود و در شاخه الگوریتم های نظارت نشده قرار دارد. برای خوشه بندی ابتدا یک بردار میانگین به صورت m_1^1, \dots, m_N^1 تعریف و مقدار دهی اولیه می کنیم که در آن N تعداد خوشه های مورد نظر است.

در ادامه مجموعه S_i که شامل نقاطی است که فاصله ی اقلیدسی آنها از خوشه i ام کوچکتر از فاصله ی آن تا سایر خوشه هاست را به صورت زیر تعریف می کنیم :

$$S_i^t = \{x_p : \|x_p - m_i^t\|^2 < \|x_p - m_j^t\|^2 \forall j, 1 \leq j \leq k\}$$

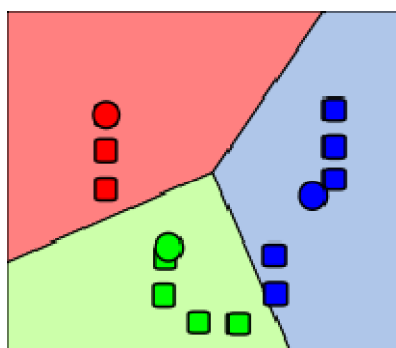
• که هر x_p باید تنها در یک مجموعه باشد ، حتی اگر شرایط عضویت در چند مجموعه را هم داشته باشد.

در ادامه در هر مرحله مرکز خوشه را به صورت زیر به روز می کنیم :

$$m_i^{t+1} = \frac{1}{|S_i^{t+1}|} \sum_{x_j \in S_i^t} x_j$$

و الگوریتم را زمانی متوقف می کنیم که تغییری در مرکز خوشه ها ایجاد نشود و در این زمان بهینه ترین حالت خوشه ها را خواهیم داشت.

- الگوریتم K-Means که توضیح داده شد ، نوع استاندارد و ابتدایی آن است و در سال های بعد از آن الگوریتم های خوشه بندی دیگری از آن مشتق شده اند که از بین آنها می توان به Fuzzy-CMeans ، K-medoids ، K-Medians اشاره کرد.[5]



شکل ۸- (مثالی از پیاده سازی K-Means) [5]



شکل-۹ (تصویر ابتدایی) [2]



شکل-۱۰ (۴ ناحیه استخراج شده از تصویر ۷ با استفاده از K-Means) [2]

۴-۵ مدل بردارها (Model Vectors):

در ادامه با محاسبه فاصله هر ناحیه با تمامی گونه های ناحیه می توان یک مدل برداری برای توصیف مفهومی محتویات بصری تصویر ایجاد کرد. برای این کار کوتاه ترین فاصله تمامی ناحیه های فریم ها را با هر گونه ی ناحیه را محاسبه می کنیم:

$$m_i(j) = \min_{r \in R(k_i)} \{d(f(w_j), f(r))\}$$

که در معادله بالا، $d(f_1, f_2) = \sqrt{f_1^2 + f_2^2}$ تابع فاصله اقلیدسی، $f(w_j)$ و $f(r)$ به ترتیب بردار ویژگی گونه ناحیه w_j و ناحیه r است. [3]

۵-۵ آنالیز پنهان مفهومی (LSA) :

مرحله بعدی برای آشکارسازی محتویات سطح بالا استفاده از تکنیک آنالیز پنهان مفهومی (LSA) است. این روش که اولین بار در مبحث پردازش زبان های طبیعی (NLP) مطرح شده است برای آشکارسازی ارتباط پنهان بین مجموعه ای از اطلاعات و اجزای تشکیل دهنده آن استفاده می شود. در این مورد فریم کلیدی به عنوان مجموعه اطلاعات و نواحی مختلف آن به عنوان اجزا در نظر گرفته می شود.

بعد از نوشتن مدل برداری m_i ، تمام مقادیر آن را نرمال کرده به طوری که در بازه $[0,1]$ قرار گیرد ، این بردار نرمال شده را با m_i' نشان می دهیم و ماتریس M را به صورت زیر تشکیل می دهیم :

$$M = \begin{pmatrix} m_1'(1) & \dots & m_{N_K}'(1) \\ \vdots & \ddots & \vdots \\ m_1'(N_T) & \dots & m_{N_K}'(N_T) \end{pmatrix}$$

- هر سطر ماتریس M که آن را با $q_i^T = (m_1'(i), \dots, m_{N_K}'(i))$ نشان می دهیم ، ارتباط بین گونه ناحیه w_i با هر فریم کلیدی k را نشان می دهد (بردار اجزا)

- هر ستون ماتریس M که آن را با $m_j' = (m_j'(1) \dots m_j'(N_T))^T$ نشان می دهیم ، ارتباط یک فریم کلیدی خاص با هر گونه ناحیه را نشان می دهد (بردار اطلاعات)

فرض کنید q_i, q_p دو بردار اجزا باشند، در این صورت ضرب داخلی آنها $q_i^T q_p$ همبستگی آنها را نشان می دهد. کاملاً واضح است که MM^T شامل تمامی این ضرب هاست. از طرف دیگر $M^T M$ نیز شامل ضرب داخلی بردارهای اطلاعات $m_i^T m_p$ است.

در ادامه ماتریس M را با روش تجزیه به مقادیر منفرد (SVD) به صورت زیر بازنویسی می کنیم:

$$M = U \Sigma V^T$$

در معادله بالا U و V دو ماتریس عمود برهم و Σ یک ماتریس قطری است:

$$M = (u_1 \dots u_{N_T}) \begin{pmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{N_T} \end{pmatrix} (v_1 \dots v_{N_T})^T$$

در معادله بالا σ_i مقادیر منفرد و v_i ، u_i بردارهای منفرد ماتریس های U و V هستند.

با کاهش ابعاد ماتریس های U ، V و Σ تا N_L ، M را می توان به صورت زیر تخمین زد:

$$\hat{M} = M_{N_L} = U_{N_L} \Sigma_{N_L} V_{N_L}^T$$

نگاشت مدل بردار m_i' به فضای مفهوم با استفاده از Σ و U در زیر آمده است:

$$\hat{m}_i = \Sigma_{N_L}^{-1} U_{N_L}^T m_i'$$

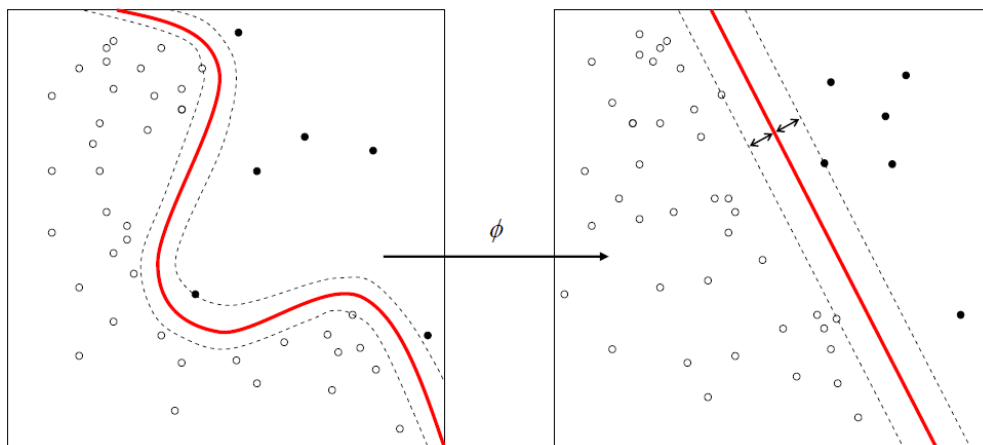
بدین ترتیب همه ی مدل های برداری استخراج شده از داده های آموزشی به فضای مفهوم نگاشت می شوند و از آنها برای آموزش مفاهیم سطح بالاتر استفاده می شود.[3]

۵-۶ استفاده از الگوریتم های دسته بندی :

در مرحله بعد برای جداپذیری نواحی در هر فضای مفهوم ، از یک الگوریتم دسته بندی غیر خطی استفاده می شود که در اینجا از SVM استفاده شده است و ورودی این شبکه مدل برداری m_i' یا خروجی الگوریتمی LSA ، \hat{m}_i' است و خروجی آن انتخاب نوع ناحیه است در صورتی که مربوط به ناحیه باشد خروجی $+1$ و در صورتی که مربوط به این منطقه نباشد خروجی -1 است ۷۰ درصد داده ها برای آموزش و ۳۰٪ آنها برای ارزیابی استفاده شده است.[3]

۵-۶-۱ ماشین بردار پشتیبانی (SVM) :

ماشین بردار پشتیبانی یکی از روش های یادگیری با نظارت است که برای طبقه بندی و رگرسیون از آن استفاده می شود ، نوع ساده این الگوریتم به صورت خطی است ولی با استفاده از ترکیب آن با Kernel Method می توان از آن برای طبقه بندی های غیر خطی نیز استفاده کرد.[5]



شکل-۱۱ (استفاده از Kernel و SVM برای طبقه بندی غیر خطی)[5]

۷-۵ استفاده از الگوریتم های فرا ابتکاری:

الگوریتم های فرا ابتکاری ، روش های بهینه سازی تقریبی هستند که توانایی حل تقریبی مسایل بهینه سازی سخت را دارند. الگوریتم های فرا ابتکاری به دسته های زیر تقسیم می شوند :

۱- مبتنی بر جواب و مبتنی بر جمعیت

۲- الهام گرفته شده از طبیعت و بدون الهام از طبیعت

۳- با حافظه و بدون حافظه

۴- قطعی و غیر قطعی

در این مقاله از الگوریتم ژنتیک ، که یک الگوریتم فرا ابتکاری مبتنی بر جمعیت است برای بهینه سازی ناحیه های پایانی تصویر استفاده شده است. [4]

۶- نتایج عملی :

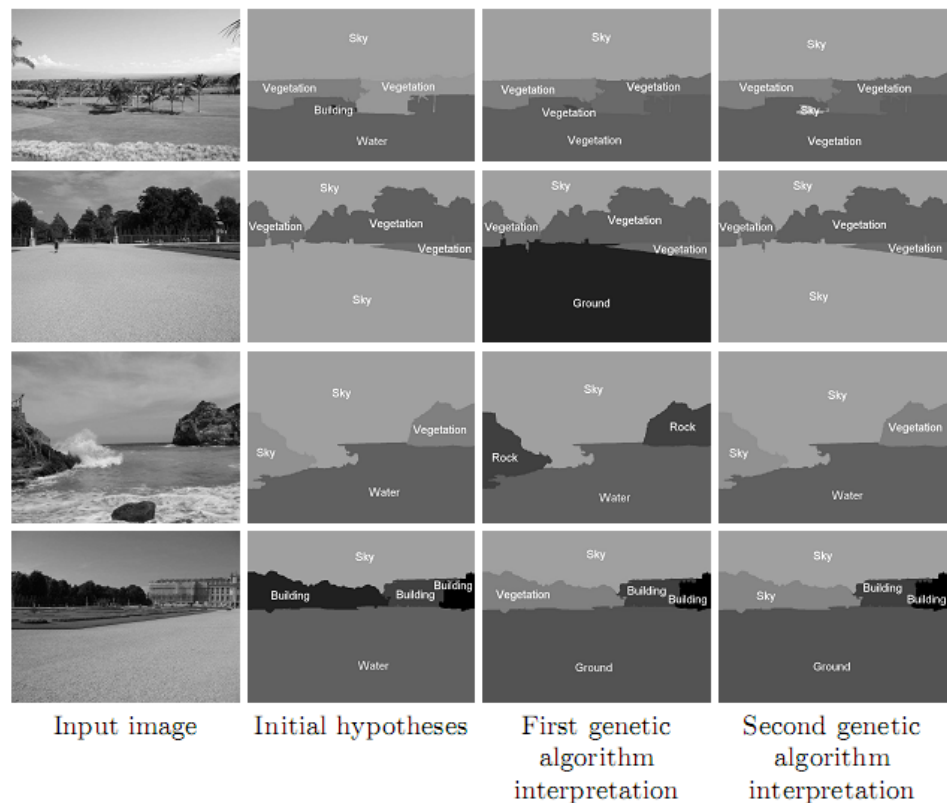
نتایج عملی این قسمت مربوط به پیاده سازی الگوریتم های بالا بر روی داده های موسسه TREC در سال ۲۰۰۷ است که شامل ۱۱۰ فیلم تقسیم بندی شده بوده و در مجموعه ۱۸۱۱۳ فریم کلیدی از آن استخراج شده است . برای هر موضوع ، مجموعه آموزش و ارزیابی جداگانه ای تشکیل شده است که از مجموعه جواب های مثبت ۷۰ درصد آن برای آموزش و ۳۰٪ دیگر برای مجموعه ارزیابی به کار رفته است و موارد منفی نیز به صورت تصادفی از سایر مجموعه فریم ها انتخاب شده اند.[3]

در جدول زیر نتایج بدون استفاده از الگوریتم ژنتیک و با در نظر گرفتن حداکثر مقدار ویژه برابر ۷۰ آمده است :

concept c_i	number of positives	$\lambda=4$						$\lambda=\max$					
		Before LSA			After LSA			Before LSA			After LSA		
		P	R	AP	P	R	AP	P	R	AP	P	R	AP
Vegetation	1939	0.643	0.312	0.460	0.626	0.221	0.395	0.322	0.313	0.232	0.268	0.222	0.179
Road	923	0.295	0.046	0.280	0.400	0.050	0.210	0.045	0.047	0.043	0.036	0.051	0.044
Explosion_Fire	29	0.291	0.777	0.182	0.200	0.111	0.148	0.000	0.000	0.001	0.001	0.111	0.000
Sky	2146	0.571	0.304	0.436	0.559	0.271	0.372	0.258	0.304	0.214	0.288	0.207	0.184
Snow	112	0.777	0.411	0.460	0.818	0.264	0.529	0.013	0.412	0.008	0.023	0.265	0.012
Office	1419	0.446	0.157	0.318	0.406	0.147	0.285	0.117	0.157	0.072	0.095	0.148	0.110
Desert	52	0.333	0.312	0.287	0.215	0.687	0.246	0.003	0.313	0.064	0.001	0.438	0.063
Outdoor	5185	0.425	0.514	0.361	0.331	0.634	0.382	0.683	0.510	0.515	0.601	0.646	0.522
Mountain	97	0.444	0.137	0.241	0.110	0.035	0.072	0.003	0.379	0.037	0.003	0.172	0.001

جدول ۱- (نتایج عملی بر روی ۹ ناحیه)[3]

در شکل ۱۰- و جدول ۲- نتایج ارزیابی با استفاده از دو الگوریتم ژنتیک و همچنین دقت و بازده آنها آمده است :



شکل ۱۲- (نتایج تشخیص موضوعی عکس) [1]

Object	Initial Hypothesis		Final interpretation 1		Final interpretation 2	
	precision	recall	precision	recall	precision	recall
Sky	93.33%	76.36%	93.33%	83.17%	93.33%	76.36%
Water	58.54%	42.86%	63.42%	52%	56.10%	46.0%
Building	81.72%	94.95%	87.83%	93.52%	81.74%	94.95%
Rock	61.11%	40.74%	83.33%	57.69%	72.22%	56.52%
Ground	56.52%	61.90%	78.26%	43.90%	69.57%	44.44%
Vegetation	42.47%	65.96%	39.73%	85.29%	41.10%	71.43%
Total precision	71.39 %		75.83 %		72.22 %	

جدول ۲- (میزان دقت و بازده الگوریتم) [1]

[1] G.Th.Papadopoulos , V.Mezaris , S.Dasiopoulou , I.Kompatsiari .: Semantic Image Analysis Using Learning Approach and Spatial Context . Semantic Multimedia , Volume4306 , 2006 , pp 199-211

[2] Al-Khatib, W., Day, Y.F., Ghafoor, A., Berra, P.B.: Semantic Annotation of Images and Videos for Multimedia Analysis. 2nd European Semantic Web Conference

[3] E. Spyrou, G. Tolias, Ph. Mylonas and Y. Avrithis

[A Semantic Multimedia Analysis Approach Utilizing a Region Thesaurus and LSA](#)

9th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2008)

[4] “Genetic Algorithm.” Wikipedia :The Free Encyclopedia. Wikimedia Foundation , Inc.2 May 2014 . Web.1 June 2014 <http://en.wikipedia.org/wiki/Genetic_algorithm>

[5] “K-Means Clustering” Wikipedia :The Free Encyclopedia. Wikimedia Foundation , Inc.3 April 2014 . Web.10 May 2014 <http://en.wikipedia.org/wiki/K-means_clustering>

[6] “Key Frame” Wikipedia :The Free Encyclopedia. Wikimedia Foundation , Inc.2 August 2013 . Web.10 May 2014 <http://en.wikipedia.org/wiki/Key_frame>

[7] Ertem Tuncel and Levent Onural, Utilization of the recursive-shortest-spanning-tree algorithm for video object segmentation by 2-D affine motion modeling, IEEE Transactions on Circuits and Systems for Video Technology, August 2000.

[8] <http://www.intechopen.com/books/image-processing/image-acquisition-storage-and-retrieval>