

## در مورد پایان‌نامه‌ی دوره‌ی کارشناسی ارشد

قبل از هر چیز، لازم می‌دانم یادآوری کنم که هر موضوعی که شما برای پایان‌نامه انتخاب می‌کنید، برای گرفتن نتیجه‌ی خوب باید در آن با جدیت تلاش کنید. بنابراین، جدا از اینکه چه موضوعی را انتخاب می‌کنید، لازم است بخش زیادی از وقتتان در هر روز را به آن اختصاص دهید: آخرین مقاله‌هایی که در آن حوزه یا موضوع‌های نزدیک به آن منتشر شده‌اند را مطالعه کنید، ایده‌های آنها را دسته‌بندی کنید و راه‌های بهبود آنها را بررسی کنید. این کار آسان نیست ولی هدف دوره‌ی کارشناسی ارشد این است که شما این توانایی را بدست آورید. شما باید بتوانید تلاش‌های خود را گزارش دهید و بیان کنید در هر هفته چه مطالعات یا آزمایش‌هایی انجام داده‌اید و چه چالش‌هایی وجود دارند و چگونه به دنبال رفع آنها هستید.

در رابطه با موضوع پایان‌نامه، هر حوزه‌ای مقدماتی دارد که برای پژوهش درست در آن، لازم است آنها را بدانید. بسیاری از درس‌های دوره‌ی کارشناسی ارشد نیز با این دید ارائه می‌شوند که مقدمات حوزه‌ای در آن مرور شوند. درس‌های پردازش موازی ترم قبل و هندسه‌ی محاسباتی ترم آینده نیز با همین دید ارائه می‌شوند. اما تفاوتی بین موضوعات تحقیقاتی و مطالبی که در درس‌ها مطالعه می‌کنید وجود دارد. پس از فراگیری مقدمات برای تحقیقات در هر زمینه‌ای، لازم است آخرین کارهای انجام شده در ارتباط با موضوع و موضوعات نزدیک به آن را مطالعه و دسته‌بندی کنید و نقاط ضعف و قوت روش‌های به کار رفته در آنها را بیابید. بنابراین، برای تعیین موضوع پایان‌نامه خواندن یک کتاب که مقدمات را بیان می‌کند کمک زیادی نمی‌کند ولی لازم است برای آشنایی با مقدمات، منابعی مثل چنین کتاب‌هایی را نیز مطالعه کنید.

## در مورد موضوع سمینار

هدف درس سمینار بررسی یک حوزه‌ی تحقیقاتی است تا با توجه به مطالعاتی که در این درس انجام می‌دهید بتوانید موضوعی را برای پایان‌نامه انتخاب کنید. بنابراین، شما در این درس با مسئله‌های مهم یک حوزه‌ی تحقیقاتی آشنا می‌شوید، روش‌های استفاده شده در حل آنها را دسته‌بندی می‌کنید و چالش‌های اصلی این حوزه را شناسایی می‌کنید. در نهایت بیان می‌کنید در کدام مسئله و برای حل کدام یک از این چالش‌ها و با چه رویکردی می‌توانید تلاش کنید.

نکته‌ی بسیار مهم دیگر، اهمیت ارائه‌ی گزارش‌های با کیفیت از مطالعاتی که انجام می‌دهید در دوره‌های تحصیلات تکمیلی است. در واقع خروجی شما همین گزارش‌ها هستند. هر مستندی را با این دید آماده کنید که قرار است افراد دیگری آن را مطالعه کنند و شما با آن مستند ارزیابی می‌شوید. یک واقعیت در تحقیقات این است که افرادی که شما را ارزیابی می‌کنند، حاضر نیستند یک متن بی‌کیفیت را مطالعه کنند (چه در ظاهر و چه از دید مفهوم). در واقع یکی از مهارت‌های مهمی که در دوره‌ی ارشد کسب می‌کنید نوشتن گزارش خوب از تحقیقات دیگران یا نتایج خودتان است. یکی از اهداف شما در دانشگاه این است که نتایجی را منتشر کنید که توسط بیشترین افراد استفاده شوند و به همین علت به دنبال با کیفیت‌ترین نتایج و انتشار آنها در معتبرترین مجله‌ها و همایش‌ها هستید.

قطعا می‌توانید در انتخاب موضوع از تزهایی که در دانشگاه‌های خوب جهان تعریف شده‌اند کمک بگیرید؛ حتی می‌توانید در سمینار خود ذکر کنید که در کدام دانشگاه، در چه دوره‌ای و در چه موضوعی تز تعریف شده است و این کار فکر بسیار خوبی است. اما دقت کنید که در درس سمینار باید دامنه‌ای فراتر از یک مسئله را مطالعه کنید. در پایان این درس می‌توانید بیان کنید: این حوزه را مطالعه کرده‌ایم که در این حوزه مسئله‌هایی که شرح داده‌ایم مطرح شده‌اند. در تحقیقات اخیر روی این مسئله‌ها، اهدافی که ذکر کرده‌ایم دنبال شده‌اند، از بین این مسئله‌ها تمرکز بیشتری روی دو مورد از آنها داشته‌ایم. رویکردهای حل این مسائل را ذکر کرده‌ایم. همچنین، چالش‌های اصلی در مسئله را شرح داده‌ایم. در مورد مسئله‌ی ...، یک تز در دوره‌ی دکترا در دانشگاه ... نیز تعریف شده است که به دنبال ... است و تز دیگری در دانشگاه ... تعریف شده است که با رویکرد ... مسئله را حل می‌کند. در نهایت با توجه به مطالعاتی که انجام داده‌ایم احتمالا می‌توانیم با استفاده از ... نتایج را بهبود دهیم.

در خاطر داشته باشید، یافتن مسئله‌های تحقیقاتی روز سخت نیست؛ مجله‌های زیادی در زمینه‌های مختلف منتشر می‌شوند و همایش‌های معتبر زیادی برگزار می‌شوند که به این مسئله‌ها می‌پردازند و مسئله‌های با اهمیت را در واقع همین مجله‌ها و همایش‌ها تعیین می‌کنند. مسئله‌ی مهم این است که در یکی از این مسئله‌ها با عمق مطالعه کنید تا بتوانید کارهای گذشته را دسته‌بندی کنید، مشکلات آنها را برشمارید و در نهایت آنها را بهبود دهید.

## فرض‌های حل مسئله

گاهی یک مسئله در شرایط و با فرض‌های متفاوتی مطالعه می‌شود؛ برخی از این شرایط در جدول زیر نشان داده شده‌اند. برای یک مسئله در هر یک از این شرایط، ممکن است الگوریتم‌های متفاوتی ارائه شده باشد و وقتی قصد بهبود الگوریتم یک مسئله‌ای را داشته باشیم، خود را به یکی از آنها محدود می‌کنیم. بنابراین در هنگام بررسی الگوریتم‌های ارائه شده برای یک مسئله، الگوریتم‌ها را با توجه به شرایط آنها ذکر کنید.

شرایط مسئله	توضیح
الگوریتم ساده	الگوریتم‌های رایج؛ هدف معمولاً کاهش پیچیدگی زمانی و حافظه است.
الگوریتم موازی	می‌تواند برای یک سخت‌افزار یا مدل موازی یا به صورت تئوری برای مدل PRAM ارائه شود؛ هدف معمولاً تسریع خوب و هزینه‌ی کم است.
الگوریتم Streaming	امکان نگهداری همه‌ی داده‌ها در حافظه وجود ندارد؛ هدف معمولاً کاهش پیچیدگی زمانی و حافظه و افزایش دقت است.
الگوریتم چند گامه (Multi-pass)	محدودیت حافظه برای نگهداری همه‌ی داده‌ها وجود دارد، ولی می‌توان ورودی را چند بار خواند. هدف معمولاً کاهش تعداد دفعات خواندن ورودی‌ها است.
الگوریتم بر خط (Online)	پس از پیش‌پردازش، هدف پاسخ به تعدادی پرسش است؛ زمان پیش‌پردازش، زمان پاسخ به هر پرسش یا حافظه‌ی مورد نیاز می‌تواند بهبود داده شود.
الگوریتم Offline	همه‌ی پرسش‌ها در دسترس هستند و می‌توانند به صورت دسته‌ای پردازش شوند. پردازش دسته‌ای می‌تواند در برخی از مسئله‌ها نسبت به الگوریتم‌های بر خط بهتر عمل کند.
الگوریتم پویا (Dynamic)	داده‌های ورودی می‌توانند تغییر کنند و پیچیدگی نگهداری ساختمان داده و پاسخ به پرسش‌ها اهمیت دارد.
الگوریتم خارجی (External)	حجم داده‌های زیاد است و داده‌ها در حافظه‌ی ثانویه ذخیره شده‌اند. در الگوریتم تعداد درخواست از حافظه‌ی ثانویه نیز تحلیل می‌شود.
الگوریتم پنجره‌ای (Time-windowed)	پرسش‌ها در یک بازه‌ی زمانی محدود می‌شوند (برای مثال نزدیک‌ترین همسایه به نقطه‌ی پرسش بین نقطه‌های موجود در زمان $t_1$ تا $t_2$ ).

## شیوهی ارائه و ارزیابی الگوریتم

برای مقایسه و ارزیابی یک الگوریتم معمولاً چند رویکرد کلی وجود دارد. گاهی الگوریتم به صورت تئوری تحلیل و مقایسه می‌شود، گاهی به صورت تجربی ارزیابی می‌گردد و گاهی این دو روش با هم ترکیب می‌شوند. این مسئله در جدول زیر به صورت خلاصه بیان شده است. اگر مسئله‌ای به صورت تجربی ارزیابی شده است، در گزارش سمینار به داده‌های استفاده شده و چگونگی انجام ارزیابی اشاره کنید.

روش ارزیابی الگوریتم	توضیح
تئوری	مزیت الگوریتم نسبت به سایر الگوریتم‌ها به صورت تئوری اثبات می‌شود. برای مثال، الگوریتمی ارائه می‌شود و نشان داده می‌شود که پیچیدگی محاسباتی آن بهتر از الگوریتم‌های پیشین است.
تجربی	الگوریتم ارائه شده با آزمایش نسبت به الگوریتم‌ها پیشین ارزیابی می‌شود. برای مثال نشان داده می‌شود که الگوریتم به صورت متوسط برای داده‌های آزمایش شده بهتر یا سریع‌تر از الگوریتم‌های مشابه عمل می‌کند. وجود داده‌های آزمایشی مناسب برای ارزیابی تجربی در این رویکرد بسیار مهم است.
ترکیبی	درستی و پیچیدگی الگوریتم به صورت تئوری اثبات می‌شود و عملکرد آن در عمل نیز ارزیابی می‌شود.

## هندسه‌ی محاسباتی در داده‌کاوی

در بررسی کاربرد هندسه‌ی محاسباتی در داده‌کاوی، به موضوع‌های زیر نیز اشاره کنید. در هر یک از این موارد، پس از بیان مسئله و کاربرد، لازم است به کارهای اصلی انجام شده در مورد آن اشاره کنید و بگویید کارهایی که اخیراً در مورد آن انجام می‌شود چه هدفی را دنبال می‌کنند. دو مورد آخر، خود مجموعه‌ای از مسئله‌ها هستند. فقط به مسئله‌های اصلی این دو اشاره کنید تا در مورد آنها توضیح دهم.

مسئله‌ی هندسی	توضیح و کاربرد
جداسازی هندسی	جدا کردن تعدادی نقطه (یا سایر اشکال هندسی) به کمک تعدادی خط (یا صفحه یا مشابه آن). کاربرد اصلی: دسته‌بندی اطلاعات. برای جداسازی هندسی تر Seara [۱] یا Vigan [۲] را مطالعه کنید.
یافتن نزدیک‌ترین همسایه‌ها	یافتن نزدیک‌ترین نقطه‌ها به نقطه‌ی پرسش. کاربرد اصلی: دسته‌بندی یک نمونه با توجه به نمونه‌های گذشته. در زمینه‌ی جستجوی نزدیک‌ترین همسایه‌ها مقاله‌های مروری زیادی وجود دارند. برای شروع فصل کتاب Nearest Neighbors in High-Dimensional Spaces نوشته شده توسط Indyk را مطالعه کنید [۳].
خوشه‌بندی هندسی	تقسیم تعدادی نقطه به دسته‌هایی که اختلاف فاصله‌ی بین نقطه‌های هر دسته (فاصله می‌تواند به شکل‌های متفاوتی تعریف شود) حداقل باشد. در زمینه‌ی خوشه‌بندی، مقاله‌ی مروری خوبی پیدا نکرده‌ام ولی می‌توانید مقدمه‌ی مقاله‌های جدیدی مثل [۴]، [۵] یا [۶] را مطالعه نمایید.
جستجوی بازه‌ای	یافتن نقطه‌هایی (یا تعداد آنها) که در یک ناحیه‌ی خاص از فضا قرار دارند. کاربرد اصلی: بازیابی اطلاعات. در زمینه‌ی جستجوی بازه‌ای مقاله‌ی کلاسیک Matousek [۷] یا فصل کتاب Agarwal [۸] را مطالعه نمایید و سپس کارهای جدیدتر را بررسی کنید.
برچسب‌گذاری	الگوریتم‌های برچسب‌گذاری نیز در نمایش نتایج داده‌کاوی کاربرد دارند. فصل پانزدهم از کتاب Tamassia [۹] را مطالعه کنید نسخه‌ی الکترونیکی این کتاب از آدرس <a href="http://cs.brown.edu/~rt/gdhandbook/">http://cs.brown.edu/~rt/gdhandbook/</a> قابل دسترس است.
سایر	در برخی از موضوعات موجود در داده‌کاوی تصویری و داده‌کاوی مکانی، مسئله‌هایی مطرح می‌شوند که با استفاده از هندسه‌ی محاسباتی می‌توان آنها را حل کرد.

در موضوع‌هایی که بررسی می‌کنید یک مصالحه وجود دارد: هر چه موضوع شناخته شده‌تر باشد، متن‌های مروری بهتری برای آن وجود دارند و بهبود نتایج اهمیت و ارزش بیشتری دارد اما با توجه به اینکه افراد بیشتری این مسئله‌ها را بررسی می‌کنند، بهبود نتایج این دسته از مسئله‌ها معمولاً سخت‌تر است.

## ساختار سمینار

تا آنجایی که من اطلاع دارم، الگوی خاصی برای شکل گزارش سمینار توسط دانشگاه تعیین نشده است. اما خوب است با تنظیمات پیشنهادی دانشگاه (در مورد فونت و شکل مراجع) برای پایان‌نامه‌ها شروع کنید؛ آن را می‌توانید از قسمت فرم‌های دانشگاه پیدا کنید.

در مورد نوع مستند، گزارش سمینار Technical Report یا Survey محسوب می‌شود. بنابراین به جای فصل، این مستند معمولاً به تعدادی بخش یا Section شکسته می‌شود. معمولاً بخش اول، مقدمه است که در آن، هدف سمینار و ساختار آن شرح داده می‌شود. بخش دوم، به مفاهیم پایه می‌پردازد. نام بخش‌های بعد، با توجه به دسته‌بندی موضوع سمینار تعیین می‌شود و بخش پایانی مربوط به نتیجه‌گیری و کارهای آتی است.

دقت کنید که گزارش سمینار نباید فقط به تعدادی مقاله اشاره کند و روش آنها را شرح دهد. در گزارش سمینار، رویکردهای حل یک مسئله به همراه مقاله‌های با اهمیت و پر تأثیر آنها دسته‌بندی می‌شوند و چالش‌های اصلی آن مطرح می‌گردند. معمولاً در قسمت‌های پایانی گزارش سمینار، جدول‌هایی این اطلاعات را به صورت خلاصه نمایش می‌دهند. الگوی یک سمینار نمونه در جدول زیر نشان داده شده است.

عنوان بخش	هدف نمونه
مقدمه	معرفی و اهمیت مسئله؛ تاریخچه‌ی کوتاه؛ بیان دقیق‌تر مسئله؛ فرضیات کلی و دامنه گزارش؛ بیان هدف هر بخش و سازماندهی گزارش (بخش دوم به ... می‌پردازد سپس بخش سوم ...).
مفاهیم پایه	مفاهیمی که در برای درک موضوع و روش‌های ارائه شده برای مسئله لازم هستند.
روش‌های مبتنی بر ...	رویکرد اول حل مسئله؛ ساختار اصلی روش‌های ارائه شده در این دسته؛ مقاله‌های مهم این دسته، تفاوت آنها و در صورت امکان دسته‌بندی آنها؛ چالش‌های مهم (برای مثال کاهش پیچیدگی حافظه با وجود بعدهای زیاد).
روش‌های مبتنی بر ...	دسته‌ی بعدی؛ مشابه بخش گذشته.
جمع‌بندی	جمع‌بندی گزارش، خلاصه‌ی دسته‌ها؛ پیشنهاد برای کارهای آتی (از جمله پایان‌نامه)، روش پیشنهادی برای بهبود و ارزیابی.

نکته‌های زیر شاید به شما در نگارش گزارش کمک کنند:

در هر رویکرد اصلی، مقاله‌های مهم یا مروری (Survey)، در صورت وجود) را بررسی کنید و سپس با دنبال کردن مقاله‌هایی که به آنها ارجاع می‌دهند (Citing articles) مجموعه‌ی مقاله‌هایتان را گسترش دهید. مقاله‌های اصلی بسیار پر اهمیت هستند چون تاریخچه‌ی مسئله را به خوبی بیان می‌کنند و روش اصلی حل مسئله را به خوبی شرح می‌دهند.

بیشتر مقاله‌هایی که در بخش‌های اصلی گزارش مطرح می‌شوند، به صورت خلاصه بیان می‌شوند. بنابراین بیشتر مقاله‌هایی را که بررسی می‌کنید لازم نیست به صورت کامل مطالعه کنید. چکیده‌ی مقاله‌ها معمولاً صورت مسئله و نتیجه را به صورت خلاصه بیان می‌کند. مطالعه‌ی چکیده، قسمت‌هایی از مقدمه و گاهی نتیجه‌گیری برای بسیاری از مقاله‌ها کافی است. ممکن است در هنگام نوشتن گزارش، قسمت‌های بیشتری از مقاله را مطالعه نمایید.

خوب است به ازای هر مقاله‌ای که مطالعه می‌کنید، فرضیات اصلی، رویکرد حل مسئله و نتیجه‌ی گزارش شده را جایی

بنویسید تا در هنگام نگارش گزارش، بتوانید به آن در بخش مناسب گزارش ارجاع دهید و در مورد آن توضیح دهید.

## در مورد تحلیل مسیر

تحلیل مسیر (Trajectory Analysis) یکی از موضوعاتی است که اخیراً در زمینه‌ی داده‌کاوی مکانی بسیار مورد توجه قرار گرفته است. برای برخی از مسئله‌های مطرح شده در این موضوع، الگوریتم‌های هندسی ارائه شده است. برای شروع تر Staals را مطالعه کنید [۱۰]. در مورد تحلیل مسیر در حالت کلی، مقاله‌ی Zheng را مطالعه کنید [۱۱] (همینطور [۱۲]). به برخی از موضوعات مطرح شده در تحلیل مسیر از دید هندسی در ادامه اشاره می‌کنم.

عنوان	توضیح
تکه تکه کردن مسیر	در حالت یکنواخت [۱۳] و غیر یکنواخت [۱۴].
یافتن ناحیه‌های داغ	مثل [۱۵].
یافتن مسیرهای مرکزی	مثل [۱۶].
گروه‌بندی مسیرها	مثل [۱۷].
مسیرهای دارای برچسب	برای شروع [۱۸] یا فصل یکم از تر Issa.

در مورد تکه تکه کردن مسیر در حالت غیر یکنواخت، ابتدا به تفاوت مسیرهای گسسته و غیر گسسته توجه کنید [۱۴]. حالت گسسته با استفاده از برنامه‌ریزی پویا با پیچیدگی زمانی و حافظه‌ی  $O(n^2)$  حل می‌شود. بهتر است تمرکز خود را روی شرط Outlier-Tolerant قرار دهید (حالتی که اختلاف مقدار در هر تکه، کمتر از مقدار  $h$  است مگر در  $\rho$  درصد مواقع). سپس الگوریتم فصل سوم را مطالعه کنید و پس از آن قسمت اول فصل ششم را مطالعه کنید. در حین مطالعه، به جزئیات الگوریتم فکر کنید: آیا مسئله با فرض متعامد بودن مسیر (مثل پنج‌شنبه‌ی بیست و ششم) آسان‌تر می‌شود؟ آیا می‌توان آن را به صورت موازی اجرا کرد؟ آیا می‌توان با حافظه‌ی بسیار کمتر (مثلاً  $O(n^{3/2})$  یا حتی  $o(n)$ ) این مسئله را حل کرد (با زمان بیشتر یا به صورت تقریبی یا احتمالی)؟

اگر نکات مبهم زیادی در الگوریتم‌های مطرح شده در [۱۴] برای شما وجود دارد، نگاهی به حالت یکنواخت این مسئله یعنی [۱۳] بیندازید.



## مجموعه‌ی مسیرها

مجموعه‌های داده‌ای در اینترنت وجود دارند که مسیر حرکت موجوداتی را بیان می‌کنند. برخی از این مجموعه‌ها در جدول زیر نمایش داده می‌شوند. لازم است در مورد هر یک از این مجموعه‌های داده (و مجموعه‌های مشابه) اطلاعاتی مثل حجم، تعداد مسیرها، متوسط تعداد نقطه‌های مسیرها، کوچک‌ترین مستطیلی که همه‌ی مسیرها در آن قرار می‌گیرند، نوع مؤلفه‌های نقطه‌ها (عدد صحیح یا اعشاری ممیز ثابت یا شناور) استخراج شوند.

مجموعه‌ی داده	توضیح
مجموعه‌ی T-Drive	مسیر حرکت تاکسی‌های در برخی از شهرهای آسیا.
مجموعه‌ی GeoLife	اطلاعات GPS حدود سه سال مربوط به ۱۸۲ نفر.
مجموعه‌ی Open Data	شامل داده‌هایی از جمله تاکسی‌های شهر نیویورک.
مجموعه‌ی Taxi Service Trajectories	اطلاعات مسیر حرکت تاکسی‌ها.
مجموعه‌ی Movebank	مسیر حرکت حیوانات.
مجموعه‌ی HURDAT	مسیر حرکت تندبادها.
مجموعه‌ی Transportation Modes	حالت حرکت افراد.
مجموعه‌ی The Greek Trucks	حرکت ماشین‌های سنگین یونان.
مجموعه‌ی User Check-in	مکان کاربران شبکه‌های اجتماعی.

## در مورد تکه‌تکه کردن مسیر

در مورد الگوریتم‌هایی که از فصل سوم تا Staal's مطالعه کرده‌اید فکر کنید [۱۰]. به خصوص، در مورد نکته‌های زیر با تمرکز و صبر فکر کنید. به صورت خلاصه در مورد آنچه فکر می‌کنید یادداشت بردارید و اگر مشکلی پیش رو می‌بینید آن را هم بنویسید؛ برای مثال، «برای پاسخ به این سؤال باید مقاله‌ی ... را مطالعه کنم»، «به نظر می‌رسد یافتن الگوریتم کارایی برای ... سخت باشد چون ...»، «اگر بتوانم قسمت ... الگوریتم را به صورت کارا انجام دهم (یا ساختمان داده‌ی مناسبی برای آن پیدا کنم)، می‌توانم این مسئله را حل کنم» یا «اطلاعات من برای پاسخ به این سؤال کافی نیست و نمی‌دانم برای پاسخ به آن چه منبعی را مطالعه کنم».

برای هر دو حالت گسسته و غیر گسسته، با فرض متعامد بودن مسیر (همواره موازی با یکی از دو محور مختصات)، آیا می‌توانید الگوریتم را ساده‌تر کنید؟ اگر خیر، در حالت یک بعدی چه طور (جسم فقط در روی یک محور مختصات حرکت کند)؟ آیا الگوریتم ساده‌ی دیگری را برای این حالت‌ها می‌توانید طراحی کنید.

آیا مسئله‌ی جالبی در گسترش الگوریتم به فضای سه بعدی به وجود می‌آید؟ به نظر می‌رسد این کار بدون تغییر الگوریتم ارائه شده ممکن باشد که خیلی جالب نیست.

اگر شرط مسئله این باشد چه طور: مسیر باید به تکه‌هایی شکسته شود که هر تکه، در یک مربع با اندازه‌ی داده شده قرار داشته باشد (مشابه ناحیه‌های داغ) به غیر از درصدی از مواقع. این درصد و اندازه‌ی مربع به عنوان ورودی داده می‌شوند. آیا این فرض در شرایطی که مقاله برای حالت کارا عنوان می‌کند می‌گنجد؟

زمان یا حافظه‌ی  $\Theta(n^2)$  در عمل بسیار زیاد است. آیا با به کار گیری ایده‌های استفاده شده در حالت یکنواخت و تغییر آن، می‌توان الگوریتمی ارائه داد که در  $O(n)$  و با حافظه‌ی کمتر برای این مسئله راه حل قابل قبولی ارائه دهد؟ مثلاً تضمین کند که جواب هیچ‌گاه بدتر از ضریبی از جواب بهینه نیست به صورت احتمالی جواب خوبی را بر می‌گرداند. فکر می‌کنم برای این کار لازم است کمی مطالعه کنید و ببینید چگونه این کار انجام می‌شود. برای نمونه [این الگوریتم پنج‌شنبه‌ی بیست و ششم](#) را مطالعه کنید.

اگر مکان یکی از نقطه‌های ورودی تغییر کند یا یک نقطه به انتهای مسیر اضافه شود، چقدر پردازش لازم است تا جواب برای حالت جدید به روز شود؟ این مسئله مهم است چون اطلاعات مسیر به تدریج به روز می‌شوند.

آیا می‌توان قسمتی از الگوریتم را به خوبی به صورت موازی اجرا کرد؟ توازی الگوریتم‌های برنامه‌ریزی پویا ساده است و کمتر اهمیت دارد؛ روی سایر قسمت‌ها تمرکز کنید.

سعی کنید با فکر کردن به سؤال‌های بالا، به دانش خود در مورد این الگوریتم عمق ببخشید و در فکر کردن عجله نکنید و چالش‌ها را

شناسایی کنید. همچنین، شاید مفید باشد فکر کنید اگر قرار باشد مسئله‌ای برای پنج‌شنبه‌های سخت در مورد این کاربرد انتخاب شود که پیاده‌سازی آن سخت نباشد، چه مسئله‌ای را انتخاب می‌کنید. تجربه نشان داده است که وقتی از دید پیاده‌سازی به یک مسئله فکر می‌کنید، جزئیاتی بیشتری از آن را کشف خواهید کرد.

پژوهش اغلب قابل پیش‌بینی نیست و معمولاً تعداد بسیار کمی از مسئله‌هایی که در مورد آنها مطالعه می‌کنید به نتیجه‌ی جالبی ختم می‌شوند. این مورد حتی در مورد مطالعات تجربی (پیاده‌سازی، مقایسه و بهبود الگوریتم‌ها) هم معمولاً صدق می‌کند. باید با شکیبایی و تمرکز مطالعه کنید و در مورد مسئله‌ها فکر کنید.

زمان	هدف
بدون زمان	مقاله‌ی Aronov و همکارانش [۱۴] در مورد گسستن مسیر غیر یکنواخت (مطالعه‌ی مجدد). کل مقاله با اهمیت است؛ حالت گسسته و پیوسته.
هفته‌ی چهارم آبان	مقاله‌ی Zheng در مورد کاربردها [۱۱]. توجه به کاربردها و مسئله‌ها؛ به خصوص بخش‌های اول، دوم و سوم، ششم، هفتم و دهم مطالعه شوند. فصل چهارم، پنجم، هشتم و نهم از اولویت پایین‌تری برخوردارند.
هفته‌ی اول آذر	مقاله‌ی Bannister و همکارانش برای پرسش‌های پنجره‌ای [۱۹]. به ایده‌ها و ساختمان‌های داده‌ی استفاده شده دقت شود.
هفته‌ی اول آذر	فصل یکم از تر Issa در مورد ترکیب مسیرها با اطلاعات محیطی.
هفته‌ی دوم آذر	مقاله‌ی Buchin و همکارانش در مورد گسستن مسیر [۱۳]. برای حالت یکنواخت ولی پیوسته.
هفته‌ی دوم آذر	به این پرسش برای حالت گسسته پاسخ دهید: استفاده از الگوریتم حریصانه‌ی حالت یکنواخت در بدترین حالت برای حالت غیر یکنواخت چه تعداد تکه‌ی اضافه ایجاد می‌کند؟
هفته‌ی سوم آذر	دریافت و بررسی مجموعه‌ی داده‌های T-drive و GeoLife؛ به ویژه سعی کنید به پرسش‌هایی که در بخش «مجموعه‌ی مسیرها» در همین مستند بیان شده است، پاسخ دهید.
هفته‌ی سوم آذر	سعی کنید الگوریتم حریصانه‌ی گسستن مسیر گسسته را برای مجموعه‌های داده‌ای که بررسی کرده‌اید پیاده‌سازی کنید. فرض کنید اختلاف سرعت در هر تکه‌ی ایجاد شده نباید بیشتر از مقداری که به عنوان ورودی داده می‌شود باشد. دقت کنید که این شرط یکنواخت است.

- سعی می‌کنم موضوعات جنبی جالب در مورد هندسه‌ی محاسباتی را نیز پیشنهاد دهم که در هنگام فراغت آنها را بررسی کنید:
- در گوگل به دنبال عبارت Computational Geometry Research Groups بگردید؛ به چند نکته دقت کنید. بیشتر دانشگاه‌های معتبر قاره‌ی آمریکا و اروپا گروه یا استنادانی دارند که در هندسه‌ی محاسباتی کار می‌کنند. اگر علاقمند بودید، ببینید روی چه موضوعاتی کار می‌کنند یا روی چه پروژه‌های تجاری کار می‌کنند.
  - به صفحه‌ی مجله‌ها و کنفرانس‌هایی که در این فایل اشاره کرده‌ام مراجعه کنید. فهرست موضوعاتی که در حوزه‌ی (Scope) آنها قرار می‌گیرد را مطالعه کنید یا نگاهی به مقاله‌های اخیر آنها بیندازید. هندسه‌ی محاسباتی در حوزه‌ی بیشتر مجله‌های دیگر کامپیوتری قرار دارد؛ صفحه‌ی ویکیپدیای هندسه‌ی محاسباتی را مشاهده کنید و حوزه‌ی این مجله‌ها را نیز بررسی کنید.
  - Lionov Wiratma دانشجوی دکترای جدید آقای van Kreveld است که روی تحلیل مسیر با توجه به محیط (Context) با رویکرد هندسه‌ی محاسباتی کار می‌کند. می‌توانید دنبال کنید در آینده روی چه موضوعاتی کار می‌کنند.
  - به کنفرانس‌های GIS/SIGSPATIAL نگاهی بیندازید که شامل موضوعاتی مثل تحلیلی داده‌های مکانی و هندسه‌ی محاسباتی است. رقابت SIGSPATIAL CUP نیز بسیار جالب است.

## هدف از مطالعه‌ی مقاله‌ها

از شما انتظار می‌رود که با یک موضوع تحقیقاتی آشنا شوید و بتوانید به سؤال‌هایی مثل سؤال‌های زیر پاسخ دهید:

- کارهای مهمی که در زمینه‌ی شکستن مسیر انجام شده‌اند چه هستند؟
- چه رویکردهای کلی برای شکستن مسیر استفاده شده‌اند؟
- در هر دسته، از چه روش‌هایی استفاده شده‌اند؟
- این کارها با چه روشی و با استفاده از چه داده‌هایی به ارزیابی و مقایسه‌ی روش‌شان پرداخته‌اند؟
- با توجه به چه پارامترهایی (مثل سرعت) مسیر شکسته شده است؟
- چالش‌ها و اشکالات در هر دسته چه بوده‌اند؟

بنابراین به ازای هر مقاله شما باید بتوانید پاسخ دهید:

- هدف مقاله چه بوده است؟
- هدف و روش این مقاله در چه دسته‌ای قرار می‌گیرد (مثال برای هدف: شکستن مسیر یا برچسب‌گذاری هر قسمت از مسیر؛ مثال برای روش: فراابتکاری یا هندسی)؟
- شکستن مسیر با چه هدفی انجام شده است؟
- شکستن مسیر با توجه به چه پارامتری (مثل سرعت، جهت، تغییر زاویه) یا چه تعریفی (مثل زمان گذر اول) انجام شده است؟
- چگونه مسیر شکسته شده است (روش مقاله)؟
- داده‌های ارزیابی این مقاله چه بوده‌اند و چگونه استفاده شده‌اند؟
- این مقاله به چه مقاله‌های مهم دیگری ارجاع داده است؟
- چه مقاله‌های مهم دیگری به این مقاله ارجاع داده‌اند؟
- حجم داده‌های آزمایش شده چقدر بوده است و آیا سرعت اجرای روش در این مقاله مورد توجه بوده است؟
- چه مقاله‌های مهم دیگری را لازم است پس از این مقاله مطالعه کنید.

لازم است بتوانید مقاله‌های مهم را تشخیص دهید. در تشخیص اهمیت یک مقاله، متغیرهای زیر کمک می‌کنند:

- در چه مجله یا کنفرانسی ارائه شده است؟ به تدریج با مجله‌ها و کنفرانس‌های معتبر حوزه آشنا خواهید شد.
- توسط چه افرادی نوشته شده‌اند؟ به تدریج با افراد تأثیرگذار در یک موضوع آشنا خواهید شد.
- چند نفر به آن مقاله ارجاع داده‌اند؟
- کارهای مرتبط را به چه دقتی بررسی کرده‌اند؟

همچنین لازم است به خوبی از سایت‌های زیر برای جستجو در مورد یک موضوع و یافتن مقاله‌هایی که به یک مقاله ارجاع داده‌اند،

استفاده کنید: <http://semanticscholar.org/> و <http://scholar.google.com/>.

بهتر است با همین مقاله‌ای که عنوانش را بیان کردید مثال بزنم. مسئله‌ی این مقاله تخمین پوشش و بهینه‌سازی شبکه‌ی بی‌سیم است؛ با توجه به مقاله، کارهای زیادی در این زمینه انجام شده است که از نمودار ورونویی و مثلث‌بندی دلانی استفاده کرده‌اند و این مقاله اطلاعات محیطی جغرافیایی را نیز، مثل برخی دیگر از مقاله‌ها لحاظ می‌کند.

این مقاله نتیجه‌ی تحقیقات عده‌ای است. اگر شما بخواهید روی زمینه‌ی مشابهی کار کنید باید دامنه‌ی مسئله‌تان را مشخص کنید. اگر قصد دارید روی شبکه‌های بی‌سیم کار کنید، لازم است اطلاعات اولیه‌ای در این زمینه بدست آورید. سپس باید بتوانید قبل از اینکه شروع به کار کنید به پرسش‌هایی پاسخ دهید:

- نوآوری‌های شما از چه جنبه‌ای خواهد بود؟
- آیا می‌خواهید روش دیگری را برای تخمین پوشش ارائه دهید؟
- چه کارهایی در این زمینه انجام شده‌اند؟
- آیا می‌خواهید از مفهومی به جز نمودار ورونویی استفاده کنید؟
- آیا می‌خواهید نمودار ورونویی را به شکلی تغییر دهید که نتیجه‌ی بهتری در تخمین پوشش بدست آورید؟
- برای نوآوری شما به چه ابزاری احتیاج خواهید داشت؟ روش‌های آماری و فراابتکاری؟
- آیا لازم است به دنبال بهبود یا حل مسئله‌ای هندسی باشید؟
- از چه دید نوآوری شما اهمیت خواهد داشت و چرا؟
- چگونه بهتر بودن نوآوری خودتان را اثبات می‌کنید؟

این پرسش‌هایی هست که از شما پرسیده می‌شود و شما باید بتوانید بعد از مطالعاتتان به آنها پاسخ دهید.

اما چیزی که من از شما انتظار دارم این است که اگر نمی‌خواهید روی موضوعات پیشنهادی من کار کنید، باید یک مسئله‌ی هندسی با اهمیت را استخراج کنید. اگر قصد دارید روی GIS کار کنید، باید مقاله‌های معتبرترین کنفرانس‌ها و مجله‌هایی را که در زمینه‌ی GIS و تحلیل داده‌های مکانی-زمانی وجود دارند بررسی کنید. سپس یک یا چند مقاله را بیابید که روی موضوعی کار می‌کنند که در آن مسئله‌ای هندسی وجود دارد. برای نمونه، بگویید که در این موضوع لازم است مسیرهای مشابه تشخیص داده شوند که در قسمتی از این موضوع، مسئله شکل هندسی پیدا می‌کند یا در این مسئله یک مسیر لازم است مسیر به تکه‌های کوچک‌تری شکسته شود که در این حالت مسئله‌ای هندسی است. حالت راحت‌تر این است که مسئله‌ی هندسی به روشنی در موضوع مطرح شده است. برای مثال، یافتن ناحیه‌ی داغ یک مسئله‌ی هندسی است که قبلاً به صورت رسمی بیان شده است یا یافتن نمودار ورونویی یک مسئله‌ی هندسی شناخته شده است.

سپس باید مشخص کنید که آیا این مسئله‌ی هندسی، قسمت پراهمیتی از موضوع مقاله‌ها است یا بهبود این قسمت تأثیر چندانی در روش حل موضوع ندارد. سپس می‌توان تصمیم گرفت که آیا بهبود عملکرد این الگوریتم هندسی، با کمک ترازوی و تقریب اهمیت دارد یا انگیزه‌ای برای بررسی داده‌های جریانی یا الگوریتم‌های یک گامه وجود دارد یا خیر یا مواردی که قبلاً به آنها اشاره کرده‌ام؛ اگر اهمیت دارد آیا قبلاً این بهبود ایجاد نشده است؟ برای مثالی که شما زده‌اید، اگر هدف‌تان بهبود الگوریتم ساختن نمودار ورونویی باشد، قطعاً ارائه‌ی الگوریتمی که به صورت موازی یا روی داده‌های جریانی کار کند بسیار پراهمیت خواهد بود از جمله در این مسئله. به عبارت دیگر شما به دنبال این هدف نیستید که یک مسئله را، با هر زمینه‌ای و به هر روشی بهبود دهید؛

چنین پژوهش‌هایی معمولاً سطحی و با ارزش پایین محسوب می‌شوند. شما به دنبال مسئله‌ی هندسی با کاربرد واقعی هستید؛ داشتن کاربرد از این دید اهمیت دارد که بهبود شما برای عده‌ای با ارزش است.

در ضمن به خاطر داشته باشید، بعد از اینکه مسئله‌ای را انتخاب کردید قسمت اصلی کارتان شروع می‌شود. حتماً دو صفحه‌ی اول همین فایل را مطالعه کنید. با توجه به این مسئله که فقط حدود یک ماه تا پایان سال باقی مانده است و با شروع کلاس‌ها زمان آزاد من بسیار کمتر خواهد بود، توصیه می‌کنم حداقل دو روز از هفته را کامل در دانشگاه حضور داشته باشید و در پایان هر دو روز جلسه‌ای با هم داشته باشیم.

1. C. Seara, *On Geometric Separability*, PhD Thesis, Universitat Politècnica De Catalunya (2002).
2. I. Vigan, *Study on Two Optimization Problems: Line Cover and Maximum Genus Embedding*, PhD Thesis, The City Univeristy of New York (2015).
3. P. Indyk, "Nearest Neighbors in High-Dimensional Spaces," pp. 877–892 in *Handbook of Discrete and Computational Geometry, 2nd Edition* (2004).
4. B. Armaselu, O. Daescu, "Maximum Area Rectangle Separating Red and Blue Points," pp. 244–251 in *The Canadian Conference on Computational Geometry* (2016).
5. S. Bandyapadhyay, A. Banik, "Polynomial Time Algorithms for Bichromatic Problems," pp. 12–23 in *CALDAM* (2017).
6. S. Har-Peled, M. Jones, "On Separating Points by Lines," in *CoRR* (2017).
7. J. Matousek, "Geometric Range Searching," *ACM Computing Surveys* **26**(4), pp. 421–461 (1994).
8. P. K. Agarwal, "Range Searching," pp. 809–837 in *Handbook of Discrete and Computational Geometry, 2nd Edition* (2004).
9. R. Tamassia, *Handbook on Graph Drawing and Visualization*, 2013.
10. F. Staals, *Geometric Algorithms for Trajectory Analysis*, PhD Thesis, Utrecht University (2015).
11. Y. Zheng, "Trajectory Data Mining - An Overview," *ACM Transactions on Intelligent Systems and Technology* **6**(3), pp. 29:1–29:41 (2015).
12. J. D. Mazimpaka, S. Timpf, "Trajectory Data Mining - A Review of Methods and Applications," *Journal of Spatial Information Science* **13**(1), pp. 61–99 (2016).
13. M. Buchin, A. Driemel, M. J. van Kreveld, V. Sacristán, "Segmenting Trajectories - A Framework and Algorithms Using Spatiotemporal Criteria," *Journal of Spatial Information Science* **3**(1), pp. 33–63 (2011).
14. B. Aronov, A. Driemel, M. J. van Kreveld, M. Löffler, F. Staals, "Segmentation of Trajectories on Nonmonotone Criteria," *ACM Transactions on Algorithms* **12**(2), pp. 26:1–26:28 (2016).
15. J. Gudmundsson, M. J. van Kreveld, F. Staals, "Algorithms for Hotspot Computation on Trajectory Data," pp. 134–143 in *SIGSPATIAL/GIS* (2013).
16. M. J. van Kreveld, M. Löffler, F. Staals, "Central Trajectories," in *CoRR* (2015).
17. K. Buchin, M. Buchin, M. J. van Kreveld, B. Speckmann, F. Staals, "Trajectory Grouping Structure," *Journal of Computational Geometry* **6**(1), pp. 75–98 (2015).
18. M. L. Damiani, H. Issa, R. H. Güting, F. Valdés, "Symbolic Trajectories and Application Challenges," *SIGSPATIAL Special* **7**(1), pp. 51–58 (2015).  
<http://dblp.org/rec/journals/sigspatial/DamianiIGV15>
19. M. J. Bannister, W. E. Devanny, M. T. Goodrich, J. A. Simons, L. Trott, "Windows into Geometric Events - Data Structures for Time-Windowed Querying of Temporal Point Sets," in *The Canadian Conference on Computational Geometry* (2014).