



دانشکده مهندسی

پایان نامه کارشناسی ارشد در رشته مهندسی کامپیوتر (هوش مصنوعی)

## یافتن الگو در دنباله های زیستی

نگارش:

الهام شرافت

استاد راهنما:

دکتر کورش زیارتی

شهریورماه ۱۳۹۰



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

به نام خدا

## اظہار نامہ

اینجانب الہام شرافت دانشجوی رشته‌ی مهندسی کامپیوتر گرایش هوش مصنوعی دانشکده‌ی برق و کامپیوتر اظہار می‌کنم که این پایان‌نامه حاصل پژوهش خودم بوده و در جاهایی که از منابع دیگران استفاده کرده‌ام، نشانی دقیق و مشخصات کامل آن را نوشته‌ام. همچنین اظہار می‌کنم که تحقیق و موضوع پایان‌نامه‌ام تکراری نیست و تعهد می‌نمایم که بدون مجوز دانشگاه دستاوردهای آن را منتشر ننموده و یا در اختیار غیر قرار ندهم. کلیه حقوق این اثر مطابق با آیین‌نامه مالکیت فکری و معنوی متعلق به دانشگاه شیراز است.

نام و نام خانوادگی: الہام شرافت

تاریخ و امضا: ۱۳۹۰/۶/۲۹



به نام خدا

یافتن الگو در دنباله‌های زیستی

به وسیله‌ی:

الهام شرافت

پایان نامه

ارائه شده به تحصیلات تکمیلی دانشگاه به عنوان بخشی  
از فعالیت‌های تحصیلی لازم برای اخذ درجه کارشناسی ارشد

در رشته:

مهندسی کامپیوتر - هوش مصنوعی

از دانشگاه شیراز

شیراز

جمهوری اسلامی ایران

ارزیابی شده توسط کمیته پایان نامه با درجه: عالی

دکتر کورش زیارتی، دانشیار بخش مهندسی و علوم کامپیوتر (رئیس کمیته) .....

دکتر حمید بیگی، دانشیار بخش مهندسی کامپیوتر

دکتر علی حمزه، استادیار بخش مهندسی و علوم کامپیوتر

شهریورماه ۱۳۹۰

تقدیم به

پدر، مادر، برادر و خواهر عزیزم، همچنین همسر مهربانم که همواره و در  
همه حال همراه و پشتیبان بنده بوده و هستند.

## سپاسگزاری

پیش از هر چیز از خداوند متعال، به سبب تمام الطافی که بی‌دریغ به من هدیه کرده است، سپاسگزارم. همچنین اکنون که این پایان‌نامه به اتمام رسیده است بر خود فرض می‌دانم که از استاد ارجمند، جناب آقای دکتر کورش زیارتی که در طول انجام این پروژه راهنمای اینجانب بوده‌اند و نیز اساتید گرامی جناب آقایان دکتر حمید بیگی و علی حمزه که مشاوره‌ی این پایان‌نامه را عهده‌دار بوده‌اند کمال تشکر و قدردانی را بنمایم. در پایان از اعضای خانواده‌ام و دوستان دیگری که با روی گشاده و در سطوح مختلف مرا یاری نموده‌اند نیز، صمیمانه سپاسگزارم.

## چکیده

### یافتن الگو در دنباله‌های زیستی

به وسیله‌ی:

#### الهام شرافت

یافتن الگوها در یک مجموعه از دنباله‌های زیستی، یکی از مسائل مهم در بیوانفورماتیک می‌باشد که گام مهمی در شناخت چگونگی تنظیم بیان ژن‌ها است. علاوه بر این می‌توان با توجه به این الگوها، اطلاعاتی راجع به کارکرد و ساختار دنباله‌های زیستی کسب کرد و به روابط تکاملی بین دنباله‌ها پی برد. حل این مسئله در پیشگویی و بررسی نواحی تنظیمی در دنباله‌ها، برای درک درست و دقیق شیوه کنترل فرآیندهای سلولی اهمیت فراوانی دارد. تا کنون الگوریتم‌های زیادی برای مسئله یافتن الگو بر اساس روش‌های تقریبی و تکراری طراحی شده‌اند، اما با این وجود نتایج حاصل از آن‌ها چندان مطلوب نیست و این مسئله همچنان نیاز به بررسی بیشتری دارد.

در این پایان‌نامه نخست مسئله یافتن الگو و مفاهیم کلیدی مربوط به آن معرفی می‌شود. سپس الگوریتم‌های معروف که برای حل مسئله یافتن موتیف ارائه شده‌اند، مورد بررسی قرار می‌گیرند. در ادامه برای نخستین بار یک الگوریتم مبتنی بر الگوریتم رقابت استعماری برای یافتن الگوها در دنباله‌های زیستی ارائه شده و نحوه‌ی کد کردن و تعریف مناسب برای عملگرهای الگوریتم و نیز تابع برازندگی جهت حل مسئله یافتن موتیف بیان می‌شود. سپس الگوریتم پیشنهادی بر روی تعدادی از داده‌های زیستی آزمایش و نتایج آن بر پایه‌ی معیار کارایی با چند الگوریتم دیگر یافتن موتیف مقایسه شده و برتری آن با استفاده از آزمون‌های آماری نشان داده می‌شود. همچنین حساسیت الگوریتم نسبت به تغییر پارامترهایش بررسی و تحلیل می‌شود.

**واژه‌های کلیدی:** بیوانفورماتیک، یافتن الگو، موتیف، *DNA*، محاسبات تکاملی، الگوریتم رقابت

استعماری، اهمیت آماری، تابع برازندگی



## فهرست مطالب

فصل اول: مقدمه.....	۱
۱-۱- مقدمه.....	۲
۲-۱- طرح مسئله و اهداف تحقیق.....	۳
۳-۱- روش تحقیق.....	۵
۴-۱- ساختار پایان نامه.....	۵
فصل دوم: مفاهیم اولیه.....	۶
۱-۲- بیوانفورماتیک.....	۷
۲-۲- مفاهیم اولیه زیست شناسی.....	۱۰
۱-۲-۲- سلول و اجزای آن.....	۱۰
۲-۲-۲- داکسی ریبو نوکلئیک اسید.....	۱۲
۳-۲-۲- ریبو نوکلئیک اسید.....	۱۴
۴-۲-۲- پروتئین.....	۱۴
۵-۲-۲- ارتباط بین درشت مولکول های DNA، RNA و پروتئین.....	۱۵
۱-۵-۲-۲- انتقال اطلاعات از DNA به RNA (ساز و کار نسخه برداری).....	۱۵
۲-۵-۲-۲- انتقال اطلاعات از RNA به پروتئین (ساز و کار ترجمه).....	۱۷
۶-۲-۲- موتیف.....	۱۸
۳-۲- الگوریتم رقابت استعماری (ICA).....	۲۰
۱-۳-۲- ساختار الگوریتم.....	۲۱
۱-۱-۳-۲- شکل دهی امپراطوری های اولیه.....	۲۲
۲-۱-۳-۲- مدل سازی سیاست جذب.....	۲۵
۳-۱-۳-۲- جابجایی موقعیت مستعمره و استعمارگر.....	۲۷
۴-۱-۳-۲- قدرت کل یک امپراطوری.....	۲۸
۵-۱-۳-۲- رقابت استعماری.....	۲۸
۶-۱-۳-۲- سقوط امپراطوری های ضعیف.....	۳۰
۷-۱-۳-۲- همگرایی.....	۳۱
۳-۳-۲- کاربردهای الگوریتم رقابت استعماری.....	۳۲
۳-۳-۲- نقش عملگرهای الگوریتم در اکتشاف و استخراج.....	۳۳
۴-۳-۲- روش کلی ایجاد نسخه های گسسته الگوریتم رقابت استعماری.....	۳۵
۴-۲- جمع بندی.....	۳۷
فصل سوم: توصیف مسئله و ادبیات موضوع.....	۳۸

- ۳-۱- موتیف از دیدگاه علوم کامپیوتر ..... ۴۰
- ۳-۲- مدل‌های نمایش موتیف ..... ۴۲
- ۳-۲-۱- نمایش موتیف به روش رشته اجماع ..... ۴۳
- ۳-۲-۲- نمایش موتیف به روش پروفایل ..... ۴۴
- ۳-۳- روش‌های تولید دنباله‌های زمینه ..... ۴۷
- ۳-۴- روش‌های امتیازدهی به موتیف‌ها ..... ۴۹
- ۳-۵- روش‌های مقایسه الگوریتم‌ها ..... ۵۱
- ۳-۶- جمع‌بندی ..... ۵۴

### فصل چهارم: بررسی الگوریتم‌های یافتن موتیف ..... ۵۵

- ۴-۱- الگوریتم‌های یافتن موتیف با مدل نمایش رشته اجماع ..... ۵۶
- ۴-۱-۱- یافتن موتیف مبتنی بر الگو ..... ۵۶
- ۴-۱-۲- یافتن موتیف مبتنی بر نمونه ..... ۵۸
- ۴-۱-۳- یافتن موتیف‌های طولانی از روی موتیف‌های کوتاه ..... ۵۹
- ۴-۱-۴- یافتن موتیف بر اساس بزرگترین زیرگراف کامل ..... ۶۰
- ۴-۱-۵- یافتن موتیف بر اساس درخت پسوندی ..... ۶۲
- ۴-۱-۶- یافتن موتیف بر اساس درختهای پیشوندی ..... ۶۳
- ۴-۱-۷- یافتن موتیف بر اساس Random projection ..... ۶۴
- ۴-۱-۸- یافتن موتیف بر اساس الگوریتم ژنتیک ..... ۶۵
- ۴-۲- الگوریتم‌های یافتن موتیف با مدل نمایش پروفایلی ..... ۶۷
- ۴-۲-۱- یافتن موتیف بر اساس درخت‌های پسوندی ..... ۶۷
- ۴-۲-۲- یافتن موتیف بر اساس الگوریتم حریم‌صانه ..... ۶۷
- ۴-۲-۳- یافتن موتیف مبتنی بر نمونه‌برداری گیبز ..... ۶۹
- ۴-۲-۴- یافتن موتیف بر اساس بیشینه کردن انتظار ..... ۷۰
- ۴-۲-۵- یافتن موتیف بر اساس الگوریتم‌های ژنتیکی ..... ۷۱

### فصل پنجم: الگوریتم رقابت استعماری جهت حل مسئله یافتن موتیف ..... ۷۲

- ۵-۱- الگوریتم پیشنهادی ..... ۷۴
- ۵-۲- ایجاد جمعیت اولیه ..... ۷۵
- ۵-۳- تابع برازندگی ..... ۷۶
- ۵-۴- عملگر جذب ..... ۷۸
- ۵-۵- عملگر انقلاب ..... ۸۰
- ۵-۶- انتخاب موتیف نهایی ..... ۸۱

### فصل ششم: آزمایش‌ها و نتایج ..... ۸۳

- ۶-۱- ارزیابی کارایی روش پیشنهادی بر روی پروموتورهای مخمر ..... ۸۴

۹۲	۶-۲- تأثیر تغییر پارامترها و انتخاب آنها بر روی کارایی الگوریتم.....
۹۴	فصل هفتم: نتیجه‌گیری و کارهای آینده
۹۷	واژه نامه فارسی به انگلیسی
۱۰۱	فهرست منابع

## فهرست جدول‌ها

فصل سوم: توصیف مسئله و ادبیات موضوع..... ۳۸

جدول ۱-۳- پرو فایل..... ۴۷

فصل ششم: آزمایش‌ها و نتایج ..... ۸۳

جدول ۱-۶- مقایسه کارایی الگوریتم‌ها بر اساس معیار nPC..... ۸۶

جدول ۲-۶- نتایج آزمون آماری ویلکسون بر روی الگوریتم ICAMF و سایر الگوریتم‌ها ۸۸

## فهرست شکل‌ها

- فصل دوم: مفاهیم اولیه** ..... ۶
- شکل ۱-۲- ساختار سلول و کروموزم ..... ۱۱
- شکل ۲-۲- زنجیره‌ای از نوکلئوتیدها و ساختار ..... ۱۳
- شکل ۳-۲- مراحل نسخه‌برداری و ترجمه ..... ۱۸
- شکل ۴-۲- فلوجارت الگوریتم رقابت استعماری ..... ۲۳
- شکل ۵-۲- چگونگی شکل‌گیری امپراطوری‌های اولیه ..... ۲۵
- شکل ۶-۲- شمای کلی حرکت مستعمرات به سمت استعمارگر ..... ۲۵
- شکل ۷-۲- حرکت مستعمرات به سمت استعمارگر در مسیر منحرف شده ..... ۲۷
- شکل ۸-۲- جابجایی موقعیت مستعمره و استعمارگر ..... ۲۸
- شکل ۹-۲- شمای کلی رقابت استعماری ..... ۲۹
- شکل ۱۰-۲- سقوط امپراطوری ضعیف (امپراطوری شماره ۴) ..... ۳۱
- شکل ۱۱-۲- شمای کلی الگوریتم رقابت استعماری ..... ۳۲
- فصل سوم: توصیف مسئله و ادبیات موضوع** ..... ۳۸
- شکل ۱-۳- چند دنباله نمونه و مصداق‌های یک موتیف در آن‌ها ..... ۴۱
- شکل ۲-۳- معنای هر یک از نمادهای آیوپاک ..... ۴۴
- فصل پنجم: الگوریتم رقابت استعماری جهت حل مسئله یافتن موتیف** ..... ۷۲
- شکل ۱-۵- گام‌های الگوریتم رقابت استعماری برای یافتن الگو ..... ۷۴
- شکل ۲-۵- گام‌های محاسبه برازش ..... ۷۸
- شکل ۳-۵- رابطه بین موقعیت شروع دنباله‌ها و عدم تطابق آن‌ها ..... ۷۹
- شکل ۴-۵- گام‌های عملگر انقلاب ..... ۸۱
- فصل ششم: آزمایش‌ها و نتایج** ..... ۸۳
- شکل ۱-۶- معیار nPC برای نتایج بدست آمده از چهار الگوریتم بر روی SCPD به روش میانگین ..... ۸۷
- شکل ۲-۶- معیار nPC برای نتایج بدست آمده از چهار الگوریتم بر روی SCPD به روش نرمالایز ..... ۸۷
- شکل ۳-۶- نتایج آزمایش دوم بر روی مجموعه داده ی GCR1 ..... ۸۹

- شکل ۶-۴- نتایج آزمایش دوم بر روی مجموعه داده ی PHO4 ..... ۸۹
- شکل ۶-۵- نتایج آزمایش دوم بر روی مجموعه داده ی RAP1 ..... ۹۰
- شکل ۶-۶- نتایج آزمایش سوم بر روی مجموعه داده ی GCR1 ..... ۹۰
- شکل ۶-۷- نتایج آزمایش سوم بر روی مجموعه داده ی PHO4 ..... ۹۱
- شکل ۶-۸- نتایج آزمایش سوم بر روی مجموعه داده ی RAP1 ..... ۹۱
- شکل ۶-۹- نتایج آزمایش سوم بر روی مجموعه داده ی STE12 ..... ۹۲
- شکل ۶-۱۰- تغییرات کارایی بر حسب تعداد جمعیت و نرخ انقلاب ..... ۹۲

## فصل اول

## ۱- مقدمه

### ۱-۱- مقدمه

بیوانفورماتیک دانش نوینی است که در آن با استفاده از کامپیوتر، نرم‌افزارهای کامپیوتری و فناوری پایگاه‌داده‌ها سعی می‌گردد تا به مسائل زیستی ویژه‌ای در زمینه‌های سلولی و ملکولی پاسخ داده شود [۱]. در این دانش با بکارگیری کامپیوتر سعی می‌گردد تا پژوهش‌های گسترده‌تری درباره ژن‌ها و پروتئین‌ها به‌عمل‌آید. دو فعالیت برجسته‌ای که دانشمندان و پژوهشگران بیوانفورماتیک به آن مشغول هستند ژنومیک<sup>۱</sup> و پروتئومیک<sup>۲</sup> هستند. ژنومیک شامل تجزیه و تحلیل داده‌ها و اطلاعات ژنتیکی بخصوص در ژنوم<sup>۳</sup> موجودات است. در حقیقت ژنوم را باید دنباله کل DNA<sup>۴</sup> موجود در سلول‌های یک جاندار دانست که به‌عنوان ماده ژنتیکی عمل می‌نماید و سبب بروز صفات وراثتی<sup>۵</sup> می‌گردد. به‌طور خلاصه باید گفت که ژنومیک شامل دنباله‌یابی<sup>۶</sup> و تحلیل ژن‌ها<sup>۷</sup> و رونوشت‌های آن‌ها در یک موجود زنده است. پروتئومیک نیز عبارتست از تحلیل پروتئین‌های یک موجود زنده.

دانش بیوانفورماتیک می‌تواند ابزاری در جهت توسعه‌ی فناوری مهندسی ژنتیک و مهندسی پروتئین باشد. برخی از پژوهشگران امروزی، فصل جدیدی در حوزه دانش بیوانفورماتیک معرفی می‌کنند که نام آن را زیست‌سیستم نهاده‌اند. زیست‌سیستم دست‌یافته‌ی جدیدی است که برای پاسخگویی به مباحث پیچیده‌ی زیستی توسط پژوهشگران به کار می‌رود. زیست‌سیستم شامل

---

<sup>1</sup> Genomics

<sup>2</sup> Proteomics

<sup>3</sup> Genome

<sup>4</sup> Deoxyribo Nucleotide Acid

<sup>5</sup> Phenotype

<sup>6</sup> Sequencing

<sup>7</sup> Gene analysis



برهم کنش میان ژنومیک و پروتئومیک و نیز داده‌های بیوانفورماتیک است که برای درک کامل از فعالیت‌های سیستم‌های زیستی به کار می‌رود. امروزه شناخت دنباله DNA به تنهایی پاسخگوی نیاز علمی دانشمندان زیست‌شناس نیست، بلکه ساز و کار عمل ملکول‌ها و اجزای سلولی نیز به شدت مورد توجه قرار گرفته است. پژوهش‌هایی که امروزه در زمینه‌ی سلول‌های بنیادین، تمایز سلولی، بیان ژن‌ها و شیوه‌ی عملکرد آن‌ها و نیز شیوه‌ی کارکرد و همکاری کلیه‌ی اجزای سلول مثل میتوکندری و پلاست‌ها صورت پذیرفته، به عنوان داده‌های خام برای دانش بیوانفورماتیک به کار می‌روند.

علم بیوانفورماتیک دارای گستردگی زیادی است و مسائل زیادی در این علم وجود دارند که برخی از آن‌ها عبارتند از: پیش‌بینی ساختار پروتئین، جستجوی تشابه، انطباق چندگانه و ساخت فیلوژنی، تحلیل دنباله‌های ژنتیکی، یافتن ژن و نگاشت ژن، بیان ژن و کشف دارو در صنایع داروسازی. یکی از نیازمندی‌های اولیه بسیاری از روش‌های ارائه شده برای حل مسائل و مشکلات نام برده شده شناسایی و یافتن الگوها در دنباله‌های ژنتیکی می‌باشد که به عنوان موتیف شناخته می‌شوند.

موتیف در زمینه تحلیل دنباله‌های بیولوژیکی، یک الگوی عام و مشترک از مبنای DNA یا آمینو اسیدها می‌باشد که دربرگیرنده یک ویژگی محفوظ شده‌ی مشترک بین گروهی از دنباله‌های DNA یا دنباله‌های پروتئین است. موتیف DNA گاهی سیگنال نامیده می‌شود. نمونه‌های آن عبارتند از: دنباله‌های تنظیمی، مکان‌های اتصال اسکافولد و مکان‌های پیوند RNA پیغام‌رسان. نمونه‌هایی از موتیف‌های پروتئین عبارتند از: مکان‌های فعال آنزیم، حوزه‌های ساختاری و برچسب‌های محلی‌سازی بافت سلولی.

## ۱-۲- طرح مسئله و اهداف تحقیق

در این پایان‌نامه نخست برخی از الگوریتم‌های یافتن موتیف مورد بررسی قرار می‌گیرد. سپس به ارائه روش جدیدی برای پیدا کردن موتیف‌های درگیر در تنظیم بیان ژن پرداخته می‌شود. در این مسئله فرض می‌شود که تعدادی دنباله داده شده است، هدف پیدا کردن یک زیردنباله مشترک در بین آن‌هاست به طوری که بیشترین شباهت را با همه داشته باشند یا به

عبارت دیگر حفاظت شده تر<sup>۱</sup> هستند [۲]. دنباله‌های داده شده در این مسئله مجموعه‌ای از نواحی پروموتور<sup>۲</sup> تعدادی از ژن‌هاست که با هم تنظیم و بیان می‌شوند.

مسئله یافتن موتیف دارای شرایط خاصی است که بر اساس فرضیه‌هایی که پژوهشگر در تحقیق خود اعمال می‌کند تعیین می‌شود. الگوریتم‌ها با استفاده از این فرضیه‌ها طراحی می‌شوند. ساختمان داده‌ای مناسب برای آن الگوریتم‌ها برگزیده می‌شود، مدل مناسب برای نمایش موتیف انتخاب می‌گردد، ورودی‌های مسئله مشخص می‌شود و سپس پاسخ‌های این الگوریتم‌ها با داده‌های زیستی موجود مقایسه می‌گردند.

به عنوان نمونه موقعیت‌ها یا زیردنباله‌های کوتاهی وجود دارند که به تعداد زیاد و با ترکیبات مختلف در دنباله‌های DNA رخ می‌دهند و محل پیوند خوردن برخی پروتئین‌های خاص به DNA هستند. اگرچه این زیردنباله‌ها به تعداد زیاد در دنباله DNA تکرار شده‌اند اما در هر کدام از این تکرارها، خطاهایی به تعداد نامشخص و در موقعیت‌های نامعین رخ داده‌اند [۳]. مثلاً ممکن است برخی از نوکلئوتیدهای زیردنباله با نوکلئوتیدهای دیگری جایگزین شوند. شرایط خاصی که در این نمونه برای این زیردنباله‌ها فرض می‌شود، تعداد رخدادهای آن‌ها در دنباله‌های DNA و خطاهایی است که در آن‌ها رخ می‌دهد. اما وجود خطاهای فراوان و ناشناخته و البته شباهت‌های تصادفی بین زیردنباله‌های غیر مرتبط، که محل پیوند خوردن هیچ پروتئینی نمی‌باشند، باعث می‌شود پیدا کردن زیردنباله‌های مربوط تبدیل به مسئله‌ای پیچیده شود. این الگوریتم‌ها اغلب از برای نمایش موتیف از ساختمان داده‌هایی چون دنباله‌ها، ماتریس‌ها و اتوماتا استفاده می‌کنند و بر اساس توابع مختلفی الگوهای پیدا شده را امتیازدهی می‌کنند تا بتوانند بر اساس این امتیازها الگوی بهینه را پیدا کنند [۲] [۴-۶] در نهایت با بهره‌گیری از برخی تئوری‌های آمار و احتمال، میزان درستی و واقعی بودن الگوها و زیردنباله‌هایی که توسط الگوریتم اعلام شده‌اند، مورد بررسی قرار می‌گیرد [۷]. بدین منظور معمولاً الگوریتم‌ها خروجی خود را با تعدادی از دنباله‌های DNA که پیش از این موتیف‌های آن کشف شده است مورد آزمایش قرار می‌دهند و به اندازه موفقیتی که در این آزمایش بدست می‌آورند به خروجی الگوریتم برای دنباله‌های DNA با موتیف‌های ناشناخته، اعتماد می‌کنند [۸].

---

<sup>1</sup> Conserved

<sup>2</sup> Promoter

### ۱-۳- روش تحقیق

الگوریتم‌های متنوعی برای مسئله یافتن موتیف ارائه شده‌اند که بیشتر آن‌ها می‌توانند در چارچوب روش نمایش موتیف به دو دسته تقسیم شوند: الگوریتم‌های جستجوی تصادفی<sup>۱</sup> و الگوریتم‌های جستجوی ترکیبی<sup>۲</sup>. الگوریتم‌های جستجوی تصادفی الگوریتم‌هایی هستند که روش نمایش آن‌ها پروفایل و الگوریتم‌های جستجوی ترکیبی الگوریتم‌هایی هستند که روش نمایش آن‌ها رشته اجماع می‌باشد [۲] [۹]. با توجه به اینکه کارایی الگوریتم‌های موجود برای یافتن موتیف رضایت‌بخش نیست [۱۰]، پژوهشگران علاقه به ارائه‌ی روش‌های جدید برای حل این مسئله دارند. در این پایان‌نامه برای نخستین بار یک الگوریتم مبتنی بر الگوریتم رقابت استعماری که الگوریتمی نسبتاً جدید در حوزه‌ی الگوریتم‌های تکاملی می‌باشد برای یافتن الگوها در دنباله‌های زیستی ارائه شده است و عملگرهای مورد نیاز آن تعریف و یک تابع برازندگی پیشنهاد داده شده است. سپس الگوریتم پیشنهادی بر روی تعدادی از داده‌های زیستی آزمایش و نتایج آن بر پایه‌ی معیار کارایی با چند الگوریتم دیگر یافتن موتیف [۱۲-۱۵] [مقایسه شده و برتری آن با استفاده از آزمون‌های آماری نشان داده می‌شود. همچنین حساسیت الگوریتم نسبت به تغییر پارامترهایش بررسی و تحلیل می‌شود.

### ۱-۴- ساختار پایان‌نامه

ادامه‌ی ساختار این پایان‌نامه به شرح زیر می‌باشد. در فصل دوم مفاهیم اولیه مورد نیاز توضیح داده شده‌اند. در فصل سوم به تعریف دقیق مسئله یافتن موتیف، روش‌های امتیازدهی و مقایسه الگوریتم‌ها پرداخته شده است. فصل چهارم مروری بر الگوریتم‌های موجود در این حوزه می‌باشد. فصل پنجم به تشریح جزئیات روش پیشنهادی پرداخته و در نهایت نتایج حاصل از مقایسه‌ی آن با سایر الگوریتم‌ها و تأثیر تغییر پارامترهای الگوریتم بر آن در فصل ششم آمده است. فصل هفتم جمع‌بندی و نتیجه‌گیری و کارهای آتی می‌باشد.

---

<sup>1</sup> Stochastic Searching Algorithms

<sup>2</sup> Combinatorial Search Algorithms

## فصل دوم