



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
دانشکده ریاضی و علوم کامپیوتر

پایان نامه کارشناسی ارشد
گرایش سیستم‌های کامپیوتری

کاهش بعد داده‌های بزرگ مقیاس با استفاده از نگاشت
تصادفی

نگارش
سیامک دهد

استاد راهنما
دکتر عادل محمدپور

استاد مشاور
دکتر هادی زارع

دی ۱۳۹۷

صفحه فرم ارزیابی و تصویب پایان نامه - فرم تأیید اعضاء کمیته دفاع

در این صفحه فرم دفاع یا تأیید و تصویب پایان نامه موسوم به فرم کمیته دفاع- موجود در پرونده آموزشی- را قرار دهید.

نکات مهم:

- نگارش پایان نامه/رساله باید به **زبان فارسی** و بر اساس آخرین نسخه دستورالعمل و راهنمای تدوین پایان نامه های دانشگاه صنعتی امیرکبیر باشد.(دستورالعمل و راهنمای حاضر)
- رنگ جلد پایان نامه/رساله چاپی کارشناسی، کارشناسی ارشد و دکترا باید به ترتیب مشکی، طوسی و سفید رنگ باشد.
- چاپ و صحافی پایان نامه/رساله بصورت **پشت و رو(دورو)** بلامانع است و انجام آن توصیه می شود.



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

تعهدنامه اصالت اثر

تاریخ: دی ۱۳۹۷

اینجانب **سیامک دهبید** متعهد می‌شوم که مطالب مندرج در این پایان‌نامه حاصل کار پژوهشی اینجانب تحت نظارت و راهنمایی اساتید دانشگاه صنعتی امیرکبیر بوده و به دستاوردهای دیگران که در این پژوهش از آنها استفاده شده است مطابق مقررات و روال متعارف ارجاع و در فهرست منابع و مآخذ ذکر گردیده است. این پایان‌نامه قبلاً برای احراز هیچ مدرک هم‌سطح یا بالاتر ارائه نگردیده است.

در صورت اثبات تخلف در هر زمان، مدرک تحصیلی صادر شده توسط دانشگاه از درجه اعتبار ساقط بوده و دانشگاه حق پیگیری قانونی خواهد داشت.

کلیه نتایج و حقوق حاصل از این پایان‌نامه متعلق به دانشگاه صنعتی امیرکبیر می‌باشد. هرگونه استفاده از نتایج علمی و عملی، واگذاری اطلاعات به دیگران یا چاپ و تکثیر، نسخه‌برداری، ترجمه و اقتباس از این پایان‌نامه بدون موافقت کتبی دانشگاه صنعتی امیرکبیر ممنوع است. نقل مطالب با ذکر مآخذ بلامانع است.

سیامک دهبید

امضا

سپاسگزاری

با تشکر از استاد گرامی دکتر محمدپور بابت همراهی و صبر ایشان

سیامک دهب
دی ۱۳۹۷

چکیده

روش تصویر تصادفی برای کاهش بعد داده‌های بزرگ مقیاس مزایای متعددی نسبت به روش‌های دیگر کاهش بعد دارد. در این پایان‌نامه این روش برای داده‌های بزرگ مقیاس با دیگر روش‌های کاهش بعد مقایسه شده است. همچنین توانایی این روش برای داده‌های با توزیع پایدار غیر نرمال با دیگر روش‌های کاهش بعد مقایسه شده است.

واژه‌های کلیدی:

کاهش بعد، تصویر تصادفی، توزیع پایدار، داده‌های بزرگ مقیاس

فهرست مطالب

صفحه	عنوان
۱	۱ مقدمه
۲	۱-۱ مقدمه
۲	۱-۱-۱ مقدمه
۳	۲ مرور ادبیات
۴	۱-۲ مقدمه
۴	۲-۲ کاهش بعد
۴	۳-۲ کاهش بعد
۴	۴-۲ کاهش بعد
۴	۵-۲ کاهش بعد
۴	۱-۵-۲ بازگیری مراجع
۵	منابع و مراجع
۶	پیوست
۷	واژه‌نامه‌ی فارسی به انگلیسی
۹	واژه‌نامه‌ی انگلیسی به فارسی

صفحه

فهرست اشکال

شکل

صفحه

فهرست جداول

جدول

فهرست نمادها

نماد	مفهوم
\mathbb{R}^n	فضای اقلیدسی با بعد n
\mathbb{S}^n	کره n یکه بعدی
M^m	خمینه m -بعدی M
$\mathfrak{X}(M)$	جبر میدان‌های برداری هموار روی M
$\mathfrak{X}^1(M)$	مجموعه میدان‌های برداری هموار 1 یکه روی (M, g)
$\Omega^p(M)$	مجموعه p -فرمی‌های روی خمینه M
Q	اپراتور ریچی
\mathcal{R}	تانسور انحنای ریمان
ric	تانسور ریچی
L	مشتق لی
Φ	۲-فرم اساسی خمینه تماسی
∇	التصاق لوی-چویتای
Δ	لاپلاسین ناهموار
∇^*	عملگر خودالحاق صوری القا شده از التصاق لوی-چویتای
g_s	متر ساساکی
∇	التصاق لوی-چویتای وابسته به متر ساساکی
Δ	عملگر لاپلاس-بلترامی روی p -فرم‌ها

فصل اول

مقدمه

۱-۱ مقدمه

۱-۱-۱ مقدمه

فصل دوم

مرور ادبیات

۱-۲ مقدمه

عمومیت پیدا کردن داده‌های حجیم مانند داده‌های حجیم تحت وب و جریان‌های داده بزرگ در کاربردهای جدید، موجب به وجود آمدن فرصت‌های و چالش‌هایی برای مهندسی و دانشمندان شده است. [۱] برای مثال، زمانی که ماتریس داده $A \in \mathbb{R}^{n \times D}$ ابعادی در حد وب داشته باشد، عملیات ساده‌ای مانند محاسبه AA^T سخت می‌شود. برای ارائه و نگهداری داده‌های حجیم در حافظه‌ای کوچک و برای استخراج اطلاعات آماری اصلی از مجموعه‌ای از بیانی محدود، روش‌های گوناگونی نمونه‌برداری توسعه یافته است. به طور کلی روش تصویر تصادفی پایدار^۱ برای داده‌های با دم سنگین خیلی خوب کار می‌کند.

روش تصویر تصادفی پایدار، ماتریس داده‌های اولیه $A \in \mathbb{R}^{n \times D}$ را در ماتریس تصادفی $R \in \mathbb{R}^{D \times k}$ ضرب می‌کند و نتیجه ماتریس $B = AR \in \mathbb{R}^{n \times k}$ است. معمولاً درایه‌های ماتریس تصادفی R به صورت i.i.d^۲ از یک توزیع α -پایدار متقارن انتخاب می‌شوند ($0 < \alpha \leq 2$). ما می‌توانیم مشخصه‌های l_α را در A بر اساس B تخمین بزنیم. در مورد حالت l_2 مزیت توزیع تصادفی پایدار توسط لم JL^۳ برجسته شده است. لم JL بیان می‌دارد که کافی است $k = O(\frac{\log n}{\epsilon^2})$ باشد تا هم فاصله دو به دویی با نرم l_α در A را بتوان با ضریب $1 \pm \epsilon$ از روی ماتریس B تخمین زد. در تر [۱] Ping Li لمی مشابه لم JL برای $0 < \alpha < 2$ اثبات شده است. روش تصویر تصادفی پایدار به یک مسئله تخمین آماری کاهش می‌یابد برای تخمین پارامتر مقیاس برای یک توزیع پایدار α متقارن. این مسئله از این جهت مورد توجه قرار می‌گیرد زیرا ما به دنبال برآوردی می‌گردیم که هم از نظر آماری درست باشند و هم از نظر محاسباتی مقرون به صرفه. برآوردهای مختلفی را مطالعه و مقایسه کردیم. شامل میانگین حسابی، میانگین هندسی، میانگین هارمونیک، تقسیم توانی^۴ و برآوردگر حداکثر بزرگنمایی.

۲-۲ کاهش بعد

۳-۲ کاهش بعد

۴-۲ کاهش بعد

۵-۲ کاهش بعد

۱-۵-۲ بارگیری مراجع

¹Stable Random Projection²Independent and identically distributed random variables³Johnson-Lindenstrauss⁴fractional power

منابع و مراجع

- [1] Li, Ping. *Stable random projections and conditional random sampling, two sampling techniques for modern massive datasets*. Stanford, 2007.

پیوست

موضوعات مرتبط با متن گزارش پایان نامه که در یکی از گروه‌های زیر قرار می‌گیرد، در بخش پیوست‌ها آورده شوند:

۱. اثبات‌های ریاضی یا عملیات ریاضی طولانی.
۲. داده و اطلاعات نمونه (های) مورد مطالعه (Case Study) چنانچه طولانی باشد.
۳. نتایج کارهای دیگران چنانچه نیاز به تفصیل باشد.
۴. مجموعه تعاریف متغیرها و پارامترها، چنانچه طولانی بوده و در متن به انجام نرسیده باشد.

کد میپل

```
with(DifferentialGeometry):  
with(Tensor):  
DGsetup([x, y, z], M)  
frame name: M  
a := evalDG(D_x)  
D_x  
b := evalDG(-2 y z D_x+2 x D_y/z^3-D_z/z^2)
```

واژه‌نامه‌ی فارسی به انگلیسی

Automorphism خودریختی	آ
د	
Degree درجه	اسکالر Scalar
ر	ب
microprocessor ریزپردازنده	بالابر Lift
ز	پ
Submodule زیرمدول	پایا Invariant
س	ت
Character سرشت	تناظر Correspondence
ص	ث
Faithful صادقانه	ثابت‌ساز Stabilizer
ض	ج
Inner product ضرب داخلی	جایگشت Permutation
ط	چ
Loop طوقه	چند جمله‌ای Polynomial
ظ	ح
Valency ظرفیت	حاصل ضرب دکارتی Cartesian product
ع	خ

Nonadjacency عدم مجاورت

ف

Vector space فضای برداری

ک

Complete reducibility کاملاً تحویل پذیر

گ

Graph گراف

م

Permutation matrix ماتریس جایگشتی

ن

Disconnected ناهمبند

و

Invertible وارون پذیر

ه

Connected همبند

ی

Edge یال

واژه‌نامه‌ی انگلیسی به فارسی

A	Lift بالا بر
Automorphism خودریختی	M
B	Module مدول
Bijection دوسویی	N
C	Natural map نگاشت طبیعی
Cycle group گروه دوری	O
D	One to One یک به یک
Degree درجه	P
E	Permutation group گروه جایگشتی
Edge یال	Q
F	Quotient graph گراف خارج‌قسمتی
Function تابع	R
G	Reducible تحویل پذیر
Group گروه	S
H	Sequence دنباله
Homomorphism همریختی	T
I	Trivial character سرشت بدیهی
Invariant پایا	U
L	

Unique منحصر بفرد

Vector space فضای برداری

V

Abstract

This page is accurate translation from Persian abstract into English.

Key Words:

Write a 3 to 5 KeyWords is essential. Example: AUT, M.Sc., Ph. D, ..



Amirkabir University of Technology
(Tehran Polytechnic)

Department of ...

MSc Thesis

Title of Thesis

By

Name Surname

Supervisor

Dr.

Advisor

Dr.

Month & Year