دانشگاه صنعتی شریف
دانشکده مهندسی کامپیوتر

عنوان:

گزارش پروژه درسی

Course Project Report Title

اعضای گروه

حامد حامدی
حسینو حسینو
محمد محمدی

نام درس

سیستم‌های عامل

نیمسال اول ۱۴۰۱-۱۴۰۲

نام استاد درس

سارا محمدی

چکیده: ئۆگارش سیستمی ئۆار علاوه بر بخش پژوهش و آمادەسازی محتوا، مستلزم رعایت ئۆکات فنی و ئۆگارشی دقیقی است که در تەیەی یک پایانئۆامەی موفق بسیار کلیدی و مؤثر است از آئۆ جایی که بسیاری از ئۆکات فنی مانۆند قالب کلی صفحات شکل و اندازەی قلم صفحات عنۆوانۆ و غیره در تەیەی پایانئۆامەها یکسانۆ است با استفاده از نۆرمافزار حروفچینۆی زیتک ایئۆ قالب میتوانۆد برای تەیەی پایانئۆامەهای کارشناسی و کارشناسی ارشد و نۆیز رسالەی دکتری مورد استفاده قرار گیرد ایئۆ نۆوشتار به طور مختصر نۆحوەی استفاده از ایئۆ قالب را نۆشانۆ میدهد

واژەهای کلیدی: پایانئۆامه حروفچینۆی، قالب، زیپرشیئۆ

۱ مقدمه

نۆخست ایئۆ فصل یک پایانئۆامه به معرفی مسئله بیانۆ اهمیت موضوع ادبیات موضوع اهداف پژوهش و معرفی ساختار پایانئۆامه میپردازد در ایئۆ فصل نۆمونۆەی مختصری از مقدمه آورده شده است

۱-۱ تعریف مسئله

ئۆگارش یک پایانئۆامه علاوه بر بخشهای پژوهش و آمادەسازی محتوا، مستلزم رعایت ئۆکات دقیق فنی و ئۆگارشی است که در تەیەی یک پایانئۆامەی موفق بسیار کلیدی و مؤثر است از آئۆ جایی که بسیاری از ئۆکات فنی مانۆند قالب کلی صفحات شکل و اندازەی قلم صفحات عنۆوانۆ و غیره در تەیەی پایانئۆامەها یکسانۆ است میتوانۆ با ارائەی یک قالب حروفچینۆی استانۆدارد ئۆگارش پایانئۆامەها را تا حد بسیار زیادی بۆود بخشید

۱-۲ اهمیت موضوع

وجود قالب استانۆدارد برای ئۆگارش پایانئۆامه از جهات مختلف حائز اهمیت است از جمله:

- ایجاد یک ئۆواخت‌ی در قالب کلی صفحات و شکل و ئۆدازه‌ی قلم‌ها
- تسه‌یل ئۆگارش پای‌ئۆئۆامه با در اخت‌یار گذاشت‌ئۆ یک قالب اولیه
- تولید خودکار صفحات دارای بخش‌های تکراری ئۆظیر صفحات ابت‌دایی و ئۆت‌های پای‌ئۆئۆامه
- پیش‌گیری از برخی خطاهای مرسوم در ئۆگارش پای‌ئۆئۆامه

۱-۳ ادبیات موضوع

اکثر دائۆش‌گاه‌های معتبر قالب استائۆداردی برای تهیه‌ی پای‌ئۆئۆامه در اخت‌یار دائۆش‌جوی‌ئۆ خود قرار می‌دهئۆدَ ای‌ئۆ قالب‌ها عموما مبت‌ئۆی بر ئۆرم‌افزاره‌ای متداول حروف‌چی‌ئۆی ئۆظیر لاتک و مای‌کروسافت ورد هست‌ئۆدَ

لاتک[1] یک ئۆرم‌افزار متئۆباز قوی برای حروف‌چی‌ئۆی متوئۆعلمی است[۱، ۲] در ای‌ئۆ ئۆوشت‌ار از ئۆرم‌افزار حروف‌چی‌ئۆی زی‌تک[2] و افزوئۆه‌ی زی‌پرش‌ی‌ئۆ[3] استفاده شده است

۱-۴ اهداف پژوهش

دائۆش‌گاه صئۆعت‌ی شریف دست‌ورالعمل جامعی را در خصوص ئۆحوه تهیه پای‌ئۆئۆامه کارش‌ئۆاسی ارشد و رساله دکت‌ری توسط کتاب‌خائۆه مرکزی دائۆش‌گاه ارائه کرده است در ای‌ئۆ ئۆوشت‌ار سعی شده است قالب استائۆداردی برای تهیه‌ی پای‌ئۆئۆامه‌ها مبت‌ئۆی بر ئۆرم‌افزار لاتک و بر اساس دست‌ورالعمل مذکور ارائه شده و ئۆحوه‌ی استفاده از قالب به طور مخت‌صر توضیح داده شودَ ای‌ئۆ قالب می‌توائۆد برای تهیه‌ی پای‌ئۆئۆامه‌های کارش‌ئۆاسی و کارش‌ئۆاسی ارشد و همچ‌ئۆی‌ئۆ رساله‌های دکت‌ری مورد استفاده قرار گیردَ

[1] LaTeX
[2] XeTeX
[3] XePersian

۲ مفاهیم اولیه

دومین فصل پایان‌نامه به طور معمول به معرفی مفاهیمی می‌پردازد که در پایان‌نامه مورد استفاده قرار می‌گیرند. در این فصل به عنوان یک نمونه کلی در خصوص نحوه نگارش پایان‌نامه و نیز برخی نکات نگارشی به اختصار توضیح داده می‌شوند.

۲-۱ نحوه‌ی نگارش

۲-۱-۱ پروندها

پرونده‌ی اصلی پایان‌نامه‌ی شما `Seminar.tex` نام دارد. پیش از شروع به نگارش سمینار بر اساس پرونده‌ی `front/info.tex` را باز نموده و مشخصات پایان‌نامه را در آن تغییر دهید.

۲-۱-۲ عبارات ریاضی

برای درج عبارات ریاضی در داخل متن از $$ و برای درج عبارات ریاضی در یک خط مجزا از $$$$ یا محیط equation استفاده کنید. برای مثال $\sum_{k=0}^{n} \binom{n}{k} = 2^n$ در داخل متن و عبارت زیر

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n \tag{1}$$

در یک خط مجزا درج شده است. دقت کنید که تمامی عبارات ریاضی، از جمله متغیره‌ای تک‌حرفی مانند x و y باید در محیط ریاضی یعنی محصور بین دو علامت $ باشند.

۲-۱-۳ علائم ریاضی پرکاربرد

برخی علائم ریاضی پرکاربرد در زیر فهرست شده‌اند. برای مشاهده‌ی دستور معادل پرونده‌ی منبع را ببینید.

- مجموعه‌های اعداد: $\mathbb{N}, \mathbb{Z}, \mathbb{Z}^+, \mathbb{Q}, \mathbb{R}, \mathbb{C}$
- مجموعه: $\{۱, ۲, \text{بگ}\}$
- دنباله: $\langle ۱, ۲, \text{بگ} \rangle$
- سقف و کف: $\lceil x \rceil, \lfloor x \rfloor$
- اندازه و متمم: $|A|, \overline{A}$
- همنهشتی: $a \equiv ۱ \ (n$ یا $a \overset{n}{\equiv} ۱$ (پیمانه‌ی n)
- ضرب و تقسیم: $\times, \cdot, \div$
- سه‌نقطه: $۱, ۲, \ldots, n$
- کسر و ترکیب: $\frac{n}{k}, \binom{n}{k}$
- اجتماع و اشتراک: $A \cup (B \cap C)$
- عمل‌گره‌ای منطقی: $\neg p \vee (q \wedge r)$
- پیکانها: $\rightarrow, \Rightarrow, \leftarrow, \Leftarrow, \leftrightarrow, \Leftrightarrow$
- عمل‌گره‌ای مقایسه‌ای: $\neq, \leqslant, \nleqslant, \geqslant, \ngeqslant$
- عمل‌گره‌ای مجموعه‌ای: $\in, \notin, \setminus, \subset, \subseteq, \subsetneq, \supset, \supseteq, \supsetneq$
- جمع و ضرب چندتایی: $\sum_{i=۱}^{n} a_i, \prod_{i=۱}^{n} a_i$
- اجتماع و اشتراک چندتایی: $\bigcup_{i=۱}^{n} A_i, \bigcap_{i=۱}^{n} A_i$
- برخی نمادها: $\infty, \emptyset, \forall, \exists, \triangle, \angle, \ell, \equiv, \therefore$

۲-۱-۴ لیست‌ها

برای ایجاد یک لیست می‌توانید از محیط‌های «فقرات» و «شمارش» همانند زیر استفاده کنید

- مورد اول
- مورد دوم
- مورد سوم

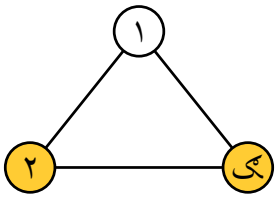
۱. مورد اول

۲. مورد دوم

ک. مورد سوم

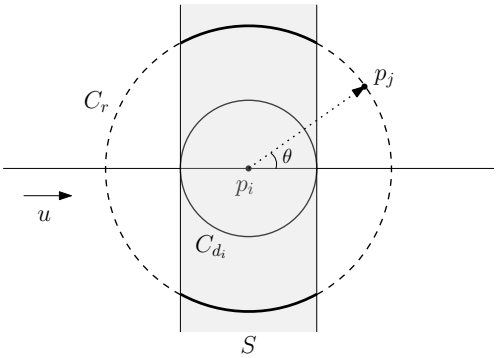
۲-۱-۵ درج شکل

یکی از روش‌های مناسب برای ایجاد شکل استفاده از نرم‌افزار LaTeX Draw و سپس درج خروجی آن به صورت یک فایل tex درون متن با استفاده از دستور fig یا centerfig است. شکل ۱ نمونه‌ای از اشکال ایجادشده با این ابزار را نشان می‌دهد.



شکل ۱: یک گراف و پوشش رأسی آن

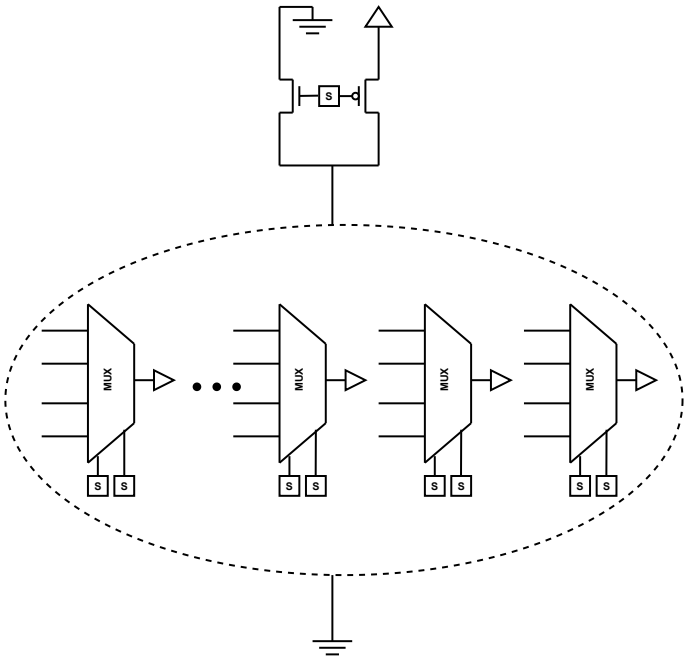
همچنین می‌توان با استفاده از نرم‌افزار Ipe شکل‌های خود را مستقیما به صورت pdf ایجاد نموده و آن‌ها را با دستورات img یا centerimg درون متن درج کنید. برای نمونه شکل ۲ را ببینید.



شکل ۲: نمونه شکل ایجادشده توسط نرم‌افزار Ipe

علاوه بر این می‌توان شکل‌ها را در ویرایشگر Drawio ایجاد نمود در این محیط امکان فراهم شده است که به صورت مستقیم به یک ویرایشگر Drawio متصل شده و شکل‌ها را ویرایش کرد به منظور کافی است بر روی فایل شکل با پسوند drawio. دوبار کلیک کرده شکل را ویرایش کنید و در انتها

آئو را ذخی‌ره ک‌ئوی‌دَ مع‌ادل pdf ای‌ئو ش‌ک‌ل در ه‌مائو ف‌ول‌دری ک‌ه ف‌ای‌ل drawio. ق‌رار دارد ذخی‌ره می‌ش‌ودَ ب‌رای ئوم‌وئوه ش‌ک‌ل ۴ را ب‌ب‌ی‌ئوی‌دَ



ش‌ک‌ل ۴: ئوم‌وئوه ش‌ک‌ل ای‌جاد ش‌ده در م‌ح‌ی‌ط Drawio. ب‌رای وی‌رای‌ش ای‌ئو ش‌ک‌ل ب‌ر روی ف‌ای‌ل test.drawio در پ‌وش‌ه figs دوب‌ار ک‌ل‌ی‌ک ک‌ئوی‌دَ

۲-۱-۶ درج ج‌دول

ب‌رای درج ج‌دول می‌ت‌وائوی‌د ب‌ا اس‌ت‌ف‌اده از دس‌ت‌ور «ج‌دول» ج‌دول را ای‌جاد ک‌رده و س‌پ‌س ب‌ا دس‌ت‌ور «ل‌وح» آئو را دروئو م‌ت‌ئو درج ک‌ئوی‌دَ ب‌رای ئوم‌وئوه ج‌دول ۱ را ب‌ب‌ی‌ئوی‌دَ

ج‌دول ۱: ع‌م‌ل‌گ‌ره‌ای م‌ق‌ای‌س‌ه‌ای

ع‌م‌ل‌گ‌ر	ع‌م‌ل‌ی‌ات
<	ک‌وچ‌ک‌ت‌ر
>	ب‌زرگ‌ت‌ر
==	م‌س‌اوی
<>	ئوام‌س‌اوی

۲-۱-۷ درج الگوریتم

برای درج الگوریتم می‌توانید از محیط «الگوریتم» همانند زیر استفاده کنید

الگوریتم ۱ پوشش رأسی حریصانه

ورودی: گراف $G = (V, E)$

خروجی: یک پوشش رأسی از G

۱: قرار بده $C = \emptyset$

۲: تا وقتی E تهی نیست:

۳: یال دلخواه $uv \in E$ را انتخاب کن

۴: رأس‌های u و v را به C اضافه کن

۵: تمام یال‌های واقع بر u یا v را از E حذف کن

۶: C را برگردان

۲-۱-۸ محیط‌های ویژه

برای درج مثال‌ها، قضیه‌ها، لم‌ها و نتیجه‌ها به ترتیب از محیط‌های «مثال»، «قضیه»، «لم» و «نتیجه» استفاده کنید برای درج اثبات قضیه‌ها و لم‌ها از محیط «اثبات» استفاده کنید

تعریف‌های داخل متن را با استفاده از دستور «مهم» به صورت **تیره** نشان دهید تعریف‌های پایه‌ای‌تر را درون محیط «تعریف» قرار دهید

تعریف ۲-۱ (اصل لانه‌کبوتری) *اگر $n+1$ کبوتر یا بیش‌تر درون n لانه قرار گیرند، آن‌گاه لانه‌ای وجود دارد که شامل حداقل دو کبوتر است.*

۲-۲ برخی نکات نگارشی

این فصل حاوی برخی نکات ابتدایی ولی بسیار مهم در نگارش متون فارسی است نکات گردآوری‌شده در این فصل به هیچ وجه کامل نیست ولی در بردارنده حداقل مواردی است که رعایت آنها در نگارش پایان‌نامه ضروری به نظر می‌رسد

۲-۲-۱ فاصله‌گذاری

۱. علائم سجاوندی مانند نقطه ویرگول دونقطه نقطه‌ویرگول علامت سؤال و علامت تعجب بدونو فاصله از کلمه‌ی پیش‌ئو خود نوشته می‌شوند ولی بعد از آنها بایدِ یک فاصله قرار گیرد مانند: من تو او

۲. علامت‌های پرانتز آکولاد کروشه نقل قول و نظایر آنها بدونو فاصله با عبارات داخل خود نوشته می‌شوند ولی با عبارات اطراف خود یک فاصله دارند مانند: (ای‌نو عبارت) یا {آنو عبارت}.

گ. دو کلمه‌ی متوالی در یک جمله همواره با یک فاصله از هم جدا می‌شوند ولی اجزای یک کلمه‌ی مرکب بایدِ با نیم‌فاصله[۴] از هم جدا شوند مانند: کتاب درس، محبت‌آمیز دوبخشی

۴. اجزای فعل‌های مرکب با فاصله از یکدیگر نوشته می‌شوند مانند: تحریر کردنو به سر آمدنو

۲-۲-۲ شکل حروف

۱. در متونو فارسی به جای حروف «ك» و «ي» عربی بایدِ از حروف «ک» و «ی» فارسی استفاده شود همچنین به جای اعداد عربی مانند ٥ و ٦ بایدِ از اعداد فارسی مانند ۵ و ۶ استفاده نمود برای این کار توصیه می‌شود صفحه‌کلیدِ فارسی استاندارد[۵] را بر روی سیستم خود نصب کنید

۲. عبارات نقل‌قول‌شده یا مؤکد بایدِ درونو علامت نقل قولِ «» قرار گیرند نه "". مانند: «کشور ایران».

گ. کسره‌ی اضافه‌ی بعد از «ه» غیرملفوظ به صورت «ه‌ی» یا «ه‌ْ» نوشته می‌شود مانند: خانه‌ی علی، دنباله‌ی فیبوناچی

---

[۴] «نیم‌فاصله» فاصله‌ای مجازی است که در عین‌نو جدا کردنو اجزای یک کلمه‌ی مرکب از یکدیگر آنها را نزدیک به هم نگه می‌دارد معمولاً برای تولید این نوع فاصله در صفحه‌کلیده‌ای استاندارد از ترکیب Shift+Sóce استفاده می‌شود

[۵] صفحه‌کلیدِ فارسی استاندارد برای ویندوز، تهیه‌شده توسط بهنام اسفند

تبصره: اگر «ه» ملفوظ باشد ئوی از به «ی» ئودارد مانند: فرمانوده دلی را پادشه خوبانو

۴. پایه‌ای همزه در کلمات همیشه «ئ» است مانند: مسئله و مسئول مگر در مواردی که همزه ساکنو است که در ایو صورت بایدمتئواسب با اعراب حرف پیش از خود ئوشته شود مانند: رأس، مؤمن

۲-۲-ک جدائوی‌سی

۱. علامت استمرار «می»، توسط ئوی‌مفاصله از جزء بعدی فعل جدا می‌شود مانند: می‌رود می‌توانوی

۲. شئواسه‌ای «ام»، «ای»، «ایم»، «اید» و «اند» توسط ئوی‌مفاصله و شئواسه‌ی «است» توسط فاصله از کلمه‌ی پیش از خود جدا می‌شوند مانند: گفته‌ام گفته‌ای، گفته است

ک. علامت جمع «ها» توسط ئوی‌مفاصله از کلمه‌ی پیش از خود جدا می‌شود مانند: ایئوها، کتابها

۴. «به» همیشه جدا از کلمه‌ی بعد از خود ئوشته می‌شود مانند: به ئوام و به آئوها، مگر در مواردی که «بـ» صفت یا فعل ساخته است مانند: بسزا، ببینوم

۵. «به» همواره با فاصله از کلمه‌ی بعد از خود ئوشته می‌شود مگر در مواردی که «به» جزئی از یک اسم یا صفت مرکب است مانند: تئواظر یکبه‌یک، سفر به تاریخ

۶. علامت صفت برتری، «تر»، و علامت صفت برتری‌ئو «تری‌ئو»، توسط ئوی‌مفاصله از کلمه‌ی پیش از خود جدا می‌شوند مانند: سئوگی‌ئوتر همه‌تری‌ئو

تبصره: کلمات «بهتر» و «بهتری‌ئو» را می‌توانو از ایئو قاعده مستثنی ئوم‌ود

ک. پیش‌ئوها و پس‌ئوهای جامد چسبیده به کلمه‌ی پیش یا پس از خود ئوشته می‌شوند مانند: همسر دانوش‌کده دانوش‌گاه

تبصره: در مواردی که خوانودئو کلمه دچار اشکال می‌شود می‌توانو پس‌ئود یا پیش‌ئود را جدا کرد مانند: همه‌ئو همارزی

۸. ضمیره‌ای متصل چسبیده به کلمه‌ی پیش از خود نوشته می‌شوند مانند: کتابم، نوامت، کلامش‌انو

۴ کاره‌ای پیشینو

در فصل سوم پایانونامه کاره‌ای پیشینو انجام‌شده روی مسئله به تفصیل توضیح داده می‌شود. نومونه‌ای از فصل کاره‌ای پیشینو در زیر آمده است[۶]

۴-۱ مسائل خوشه‌بندی

مسئله‌ی **خوشه‌بندی**[۷] یکی از مهمترینو مسائل در زمینه‌ی داده‌کاوی به حساب می‌آید. در اینو مسئله هدف دسته‌بندی تعدادی شیء به‌گونه‌ای است که اشیاء درونو یک دسته (خوشه)، نوسبت به یکدیگر در برابر دسته‌های دیگر شبیه‌تر باشند (معیاره‌ای متفاوتی برای تشابه تعریف می‌گردد). اینو مسئله در حوزه‌های مختلفی از علوم کامپیوتر از جمله داده‌کاوی، جست‌وجوی الگو[۸]، پردازش تصویر[۹]، بازیابی اطلاعات[۱۰] و رایانوش زیستی[۱۱] مورد استفاده قرار می‌گیرد [۴].

تا کنونو راه‌حل‌های زیادی برای اینو مسئله ارائه شده است که از لحاظ معیار تشخیص خوشه‌ها و نوحوه‌ی انتخاب یک خوشه با یکدیگر تفاوت بسیاری دارند. به همینو خاطر مسئله‌ی خوشه‌بندی یک مسئله‌ی بهینه‌سازی چندهدفه[۱۲] محسوب می‌شود.

همانو طور که در مرجع [۴] ذکر شده است، خوشه در خوشه‌بندی تعریف واحدی ندارد و یکی از دلایل وجود الگوریتم‌های متفاوت، همینو تفاوت تعریف‌ها از خوشه است. بنابراینو با توجه به مدلی که برای خوشه‌ها ارائه می‌شود، الگوریتم متفاوتی نیز ارائه می‌گردد. در ادامه به بررسی تعدادی از معروف‌ترینو مدل‌های مطرح می‌پردازیم:

---

۶ مطالب اینو فصل نومونه از پایانونامه‌ی آقای بهنام حاتمی گرفته شده است.

۷ Clustering

۸ Pattern recognition

۹ Image analysis

۱۰ Information retrieval

۱۱ Bioinformatics

۱۲ Multi-objective

- **مدل‌های مرکزگرا:** در این مدل‌ها، هر دسته با یک مرکز نشان داده می‌شود. از جمله معروف‌ترین روش‌های خوشه‌بندی بر اساس این مدل خوشه‌بندی k-مرکز، خوشه‌بندی k-میانگین[۱۳] و خوشه‌بندی k-میانه[۱۴] است.

- **مدل‌های مبتنی بر توزیع نقاط:** در این مدل دسته‌ها با فرض پی‌روی از یک توزیع احتمالی مشخص می‌شوند. از جمله الگوریتم‌های معروف ارائه شده در این مدل الگوریتم بیشینه‌سازی امید ریاضی[۱۵] است.

- **مدل‌های مبتنی بر تراکم نقاط:** در این مدل خوشه‌ها متناسب با تراکم نقاط در مجموعه داده مورد استفاده قرار می‌گیرد.

- **مدل‌های مبتنی بر گراف:** در این مدل هر خوشه به مجموعه از رئوس گفته می‌شود که تمام رئوس آن با یکدیگر همسایه باشند. از جمله الگوریتم‌های معروف این مدل الگوریتم خوشه‌بندی HCS[۱۶] است.

الگوریتم‌های ارائه شده تنها از نظر نوع مدل با یکدیگر متفاوت نیستند بلکه می‌توانند آنها را از لحاظ نحوه تخصیص نقاط به خوشه‌ها نیز تقسیم‌بندی کرد:

- **تخصیص قطعی داده‌ها:** در این نوع خوشه‌بندی هر داده دقیقاً به یک خوشه اختصاص داده می‌شود.

- **تخصیص قطعی داده‌ها با داده‌ی پرت:** در این نوع خوشه‌بندی ممکن است بعضی از داده‌ها به هیچ خوشه‌ای اختصاص نیابد اما بقیه داده‌ها هر کدام دقیقاً به یک خوشه اختصاص می‌یابد.

- **تخصیص قطعی داده:** در این نوع خوشه‌بندی هر داده دقیقاً به یک خوشه اختصاص داده می‌شود.

- **خوشه‌بندی همپوشان:** در این نوع خوشه‌بندی هر داده می‌تواند به چند خوشه اختصاص داده شود. در گونه‌ای از این مدل می‌توان هر نقطه را با

[۱۳] k-Means
[۱۴] k-Median
[۱۵] Expectation-maximization
[۱۶] Highly Connected Subgraphs

احتمالی به هر خوشه اختصاص می‌یابد به ایئو گوئوه از خوشه‌بئودی، خوشه‌بئودی ئورم[۱۵] گفته می‌شود

- **خوشه‌بئودی سلسه‌مراتبی:** در ایئو ئوع خوشه‌ها، داده‌ها به گوئوه‌ای به خوشه‌ها تخصیص داده می‌شود که دو خوشه یا اشتراک ئودارئود یا یکی به طور کامل دیگری را می‌پوشانئود در واقع در بی‌ئو خوشه‌ها، رابطه‌ی پدر فرزئودی برقرار است

در بی‌ئو دسته‌بئودی‌های ذکر شده تمرکز اصلی ایئو پایانئوئامه بر روی مدل مرکزگرا و خوشه‌بئودی قطعی با داده‌های پرت با مدل k-مرکز است همانطور که ذکر شد علاوه بر مسئله‌ی k-مرکز که به تفصیل مورد بررسی قرار می‌گیرد k-میانئوه و k-میانئوگی‌ئو از جمله معروف‌تری‌ئو خوشه‌بئودی‌های مدل مرکزگرا هستئود در خوشه‌بئودی k-میانئوه هدف افراز ئوقاط به k خوشه است به گوئوه‌ای که مجموع مربع فاصله‌ی هر ئوقطه از میانئوه‌ی ئوقاط آئو خوشه کمینئوه گردد در خوشه‌بئودی k-میانئوگی‌ئو هدف افراز ئوقاط به k خوشه است به گوئوه‌ای که مجموع فاصله‌ی هر ئوقطه از میانئوگی‌ئو ئوقاط داخل خوشه (یا مرکز آئو خوشه) کمینئوه گردد

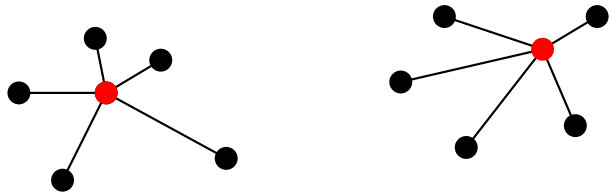
۲-۲ خوشه‌بئودی k-مرکز

یکی از رویکردهای شناخته‌شده برای مسئله‌ی خوشه‌بئودی، مسئله‌ی **k-مرکز** است در ایئو مسئله هدف پیدا کردئو k ئوقطه به عنوانئو مرکز دسته‌ها است به‌طوری‌که شعاع دسته‌ها تا حد ممکنئو کمینئوه شود مثالی از مسئله‌ی ۲-مرکز در شکل ۴ نشانئو داده شده است در ایئو پژوهش، مسئله‌ی k-مرکز با متری‌که‌ای خاص و برای kهای کوچک مورد بررسی قرار گرفته است و هر کدام از تعریف رسمی مسئله‌ی k-مرکز در زیر آمده است:

مسئله‌ی ۲-۱ **(kـ مرکز)** *گراف کامل بدون جهت $G = (V, E)$ با تابع فاصله‌ی d، که از نامساوی مثلثی پیروی می‌کند داده شده است. زیرمجموعه‌ی $S \subseteq V$ با اندازه‌ی k را به‌گونه‌ای انتخاب کنید که عبارت زیر را کمینه کند:*

$$\max_{v \in V}\{\min_{s \in S} d(v, s)\} \tag{۲}$$

[۱۵] Soft clustering



شکل ۴: نمونه‌ای از مسئله‌ی ۲-مرکز

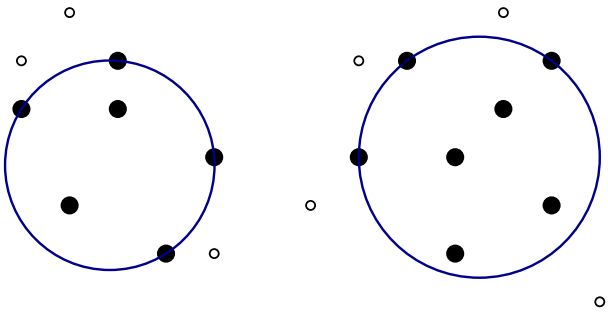
گونه‌های مختلفی از مسئله‌ی k-مرکز با محدودیت‌های متفاوت توسط پژوهشگران مورد مطالعه قرار گرفته است. از جمله‌ی این گونه‌ها، می‌توان به حالتی که در بین داده‌های ورودی، داده‌ای پرت وجود دارد اشاره کرد. در واقع در این مسئله قبل از خوشه‌بندی می‌توانیم تعدادی از نقاط ورودی را حذف نموده و سپس به خوشه‌بندی نقاط بپردازیم. سخت‌ترین این مسئله از آنجاست که نه تنها باید مسئله‌ی خوشه‌بندی را حل نمود بلکه در ابتدا باید تصمیم گرفت که کدام یک از داده‌ها را به‌عنوان داده‌ی پرت در نظر گرفت که بزرگترین جواب در زمان خوشه‌بندی به دست آید. در واقع اگر تعداد نقاط پرتی که مجاز به حذف است برابر صفر باشد مسئله به مسئله‌ی k-مرکز تبدیل می‌شود. نمونه‌ای از مسئله‌ی ۲-مرکز با داده‌ی پرت را در شکل ۵ می‌توان دید. تعریف دقیق‌تر این مسئله در زیر آمده است:

مسئله‌ی ۲-گ **(kـ مرکز با داده‌های پرت)** *یک گراف کامل بدون جهت $G = (V, E)$ با تابع فاصله‌ی d، که از نامساوی مثلثی پیروی می‌کند داده‌شده است. زیرمجموعه‌ی $Z \subseteq V$ با اندازه‌ی z و مجموعه‌ی $S \subseteq V - Z$ با اندازه‌ی k را انتخاب کنید به طوری که عبارت زیر را کمینه کند:*

$$\max_{v \in V - Z} \{\min_{s \in S} d(v, s)\} \qquad (\text{گ})$$

گونه‌ی دیگری از مسئله‌ی k-مرکز که در سال‌های اخیر مورد توجه قرار گرفته است حالت جویبار داده‌ی آن است که در این‌گونه از مسئله‌ی k-مرکز در ابتدا تمام نقاط در دسترس نقاط سنتی بلکه به‌مرور زمان نقاط در دسترس قرار می‌گیرند. محدودیت دومی که وجود دارد محدودیت حافظه است به‌طوری‌که نمی‌توان تمام نقاط را در حافظه نگه داشت و بعضاً حتی امکان نگه‌داری در حافظه‌ی جانبی نیز وجود ندارد و به‌طور معمول باید مرتبه‌ی حافظه‌ای کمتر از مرتبه حافظه‌ی **خطی**[۱۸] متناسب با تعداد نقاط استفاده نمود. از این به بعد به چنین

[۱۸] Linear



شکل ۵: نمونه‌ای از مسئله‌ی ۲-مرکز با داده‌های پرت

مرتبه‌ای، مرتبه‌ی **زیرخطی**[۱۹] می‌گویی‌م مدلی که ما در این پژوهش بر روی آن تمرکز داری‌م مدل جوی‌بار داده تک‌گذره[۲۰] [۵] است‌ِ ی‌عنی تنها یک بار می‌توان از ابتدا تا انتهای داده‌ها را بررسی کرد و پس از عبور از یک داده‌ٔ اگر آن داده در حافظه ذخیره نشده باشد دیگر به آن دست‌رسی وجود ندارد‌ِ علاوه بر این‌ٔ در هر لحظه باید بتوانیم به پرسم‌انی (برای تمام نقاطی از جوی‌بار داده که تاکنون به آن دست‌رسی داشته‌ای‌م) پاسخ داد‌ِ

مسئله‌ی مک-مک (kـ **مرکز در حالت جویبار داده**) *مجموعه‌ای از نقاط در فضای dـ بعدی به مرور زمان داده می‌شود. در هر لحظه از زمان، به ازای مجموعه‌ی U از نقاطی که تا کنون وارد شده‌اند، زیرمجموعه‌ی $S \subseteq U$ با اندازه‌ی k را انتخاب کنید به طوری‌که عبارت زیر کمینه شود:*

$$\max_{u \in U}\{\min_{s \in S} d(u, s)\} \tag{۴}$$

از آن‌جایی که گونه‌ی جوی‌بار داده و داده پرت مسئله‌ی k-مرکز به علت به‌روز بودن مباحث داده‌های حجی‌م[۲۱]، به تازگی مورد توجه قرار گرفته است‌ِ در این‌ٔ تحقیق سعی شده است که تمرکز بر روی این‌ٔگونه‌ی خاص از مسئله باشد‌ِ همچنین‌ٔ در این‌ٔ پژوهش سعی می‌شود گونه‌ای مسئله را برای انواع متری‌کها و برای kهای کوچک نیز مورد بررسی قرار داد‌ِ

[۱۹] sublinear
[۲۰] Single pass
[۲۱] Big data

۴-۴ مدل جویبار داده

همانطور که ذکر شد مسئله‌ی k-مرکز در حالت داده‌ای پرت و جویبار داده گونه‌ای تعمیم‌یافته از مسئله‌ی k-مرکز است و در حالت‌های خاص به مسئله‌ی k-مرکز کاهش پیدا می‌کنند. مسئله‌ی k-مرکز در حوزه‌ی مسائل انتخاب-سخت[۴۴] قرار می‌گیرد و با فرض $P \neq NP$ الگوریتم دقیق با زمان چندجمله‌ای برای آن وجود ندارد [۶]. بنابراین برای حل کارای[۴۲] این مسائل از الگوریتم‌های تقریبی[۲۴] استفاده می‌شود.

برای مسئله‌ی k-مرکز دو الگوریتم تقریبی معروف وجود دارد. در الگوریتم اول که به روش حریصانه[۲۵] عمل می‌کند در هر مرحله بزرگترین مرکز ممکن را انتخاب می‌کند به طوری تا حد ممکن از مراکز قبلی دور باشد [۷]. این الگوریتم الگوریتم تقریبی با ضریب تقریب ۲ ارائه می‌دهد. در الگوریتم دوم با استفاده از مسئله‌ی مجموعه‌ی غالب کمینه[۳۸]، الگوریتمی با ضریب تقریب ۲ ارائه می‌گردد [۸]. همچنین ثابت شده است که بهتر از این ضریب تقریب، الگوریتمی نمی‌تواند ارائه داد مگر آنکه $P = NP$ باشد.

برای مسئله‌ی k-مرکز در حالت جویبار داده برای ابعاد بالا، بهترین الگوریتم موجود ضریب تقریب $2 + \epsilon$ دارد [۹، ۱۰، ۱۱] و ثابت می‌شود الگوریتمی با ضریب تقریب بهتر از ۲ نمی‌تواند ارائه داد. برای مسئله‌ی k-مرکز با داده‌ی پرت در حالت جویبار داده نیز بهترین الگوریتم ارائه شده الگوریتمی با ضریب تقریب $4 + \epsilon$ است که با کرانی پایینی نزدیک همخوان اختلاف قابل توجهی دارد [۱۲].

برای kهای کوچک به خصوص $k = 1, 2$، الگوریتم‌های بهتری ارائه شده است. بهترین الگوریتم ارائه شده برای مسئله‌ی ۱-مرکز در حالت جویبار داده برای ابعاد بالا، دارای ضریب تقریب ۱/۲۲ است و کرانی پایینی $\frac{1+\sqrt{2}}{2}$ نیز برای این مسئله اثبات شده است [۱۳، ۱۴]. برای مسئله ۲-مرکز در حالت جویبار داده برای ابعاد بالا، اخیرا راه‌حلی با ضریب تقریب $1/8 + \epsilon$ ارائه شده است [۱۵]. برای مسئله‌ی ۱-مرکز با داده‌ی پرت تنها الگوریتم موجود الگوریتمی با

---

[۴۴] NP-hard
[۴۲] Efficient
[۲۴] Approximation algorithm
[۲۵] Greedy
[۳۸] Dominating set

جدول ۲: ئۆمونئوەەایی از کرائۆ پایی‌ئۆ تقریب‌پذیری مسائل خوش‌ەبئۆدی

مسئله	کرائۆ پایی‌ئۆ تقریب‌پذیری
k-مرکز	۲ [۸]
k-مرکز در فضای اقلیدسی	۱/۸۲۲ [۱۵]
۱-مرکز در حالت جویبار داده	$\frac{1+\sqrt{2}}{2}$ [۱۵]
k-مرکز با ئۆقاط پرت و ئۆقاط اجباری	۳ [۱۲]

ضریب تقریب ۳/۱ است [۱۶].

۴-۳ تقریب‌پذیری

یکی از راه‌کارەایی که برای کارآمد کردئۆ راه‌حل ارائه شده برای یک مسئله وجود دارد استفاده از الگوریتم‌ەای تقریبی برای حل آئۆ مسئله است. یکی از عمده‌ترین دغدغه‌ەای مطرح در الگوریتم‌ەای تقریبی کاهش ضریب تقریب است. در بعضی از موارد حتی امکائۆ ارائه‌ی الگوریتم تقریبی با ضریبی ثابت ئۆیز وجود ئۆدارد. به طور مثال الگوریتم تقریبی با ضریب تقریب کمتر از ۲، برای مسئله‌ی k-مرکز وجود ئۆدارد مگر ایئۆکه $P = NP$ باشد. برای مسائل مختلف معمولاً می‌توائۆ کرائۆ پایی‌ئۆی برای می‌زائۆ تقریب‌پذیری آئۆها ارائه داد. در واقع برای برخی مسائل ائۆپی-سخت علاوه بر ایئۆ که الگوریتم کارآمدی وجود ئۆدارد بعضاً الگوریتم تقریبی با ضریبی تقریب کم و ئۆزدیک به یک ئۆیز وجود ئۆدارد. در جدول ۲ می‌زائۆ تقریب‌پذیری مسائل مختلفی که در ایئۆ پایائۆئۆامه مورد استفاده قرار می‌گیرد را می‌بیئۆید.

References

[1] D. E. Knuth. *The TEXbook*. Addison-Wesley, 1984.

[2] L. Lamport. *LATEX—A Document Preparation System*. Addison-Wesley, 1985.

[3] J. Han and M. Kamber. *Data Mining, Southeast Asia Edition: Concepts and Techniques*. Morgan kaufmann, 2006.

[4] V. Estivill-Castro, "Why so many clustering algorithms: a position paper," *ACM SIGKDD explorations newsletter*, vol.4, no.1, pp.65–75, 2002.

[5] C. C. Aggarwal. *Data streams: models and algorithms*. Springer Science & Business Media, 2007.

[6] M. R. Garey and D. S. Johnson, "Computers and intractability: a guide to the theory of NP-completeness," *Freeman & Co.*, 1979.

[7] N. Megiddo and K. J. Supowit, "On the complexity of some common geometric location problems," *SIAM Journal on Computing*, vol.13, no.1, pp.182–196, 1984.

[8] V. V. Vazirani. *Approximation Algorithms*. Springer-Verlag New York, Inc., 2001.

[9] R. M. McCutchen and S. Khuller, "Streaming algorithms for k-center clustering with outliers and with anonymity," in *Proceedings of the 11th International Workshop on Approximation Algorithms*, pp.165–178, 2008.

[10] S. Guha, "Tight results for clustering and summarizing data streams," in *Proceedings of the 12th International Conference on Database Theory*, pp.268–275, 2009.

[11] H.-K. Ahn, H.-S. Kim, S.-S. Kim, and W. Son, "Computing k centers over streaming data for small k," *International Journal of Computational Geometry and Applications*, vol.24, no.02, pp.107–123, 2014.

[12] M. Charikar, S. Khuller, D. M. Mount, and G. Narasimhan, "Algorithms for facility location problems with outliers," in *Proceedings of the 12th ACM-SIAM Symposium on Discrete Algorithms*, pp.642–651, 2001.

[13] P. K. Agarwal and R. Sharathkumar, "Streaming algorithms for extent problems in high dimensions," in *Proceedings of the 21st ACM-SIAM Symposium on Discrete Algorithms*, pp.1481–1489, 2010.

[14] T. M. Chan and V. Pathak, "Streaming and dynamic algorithms for minimum enclosing balls in high dimensions," *Computational Geometry: Theory and Applications*, vol.47, no.2, pp.240–247, 2014.

[15] S.-S. Kim and H.-K. Ahn, "An improved data stream algorithm for clustering," in *Proceedings of the 11th Latin American Symposium on Theoretical Informatics*, pp.273–284, 2014.

[16] H. Zarrabi-Zadeh and A. Mukhopadhyay, "Streaming 1-center with outliers in high dimensions.," in *Proceedings of the 21st Canadian Conference on Computational Geometry*, pp.83–86, 2009.

[17] M. Bern and D. Eppstein, "Approximation algorithms for NP-hard problems," chap. Approximation Algorithms for Geometric Problems, pp.296–345, PWS Publishing Co., 1997.

آ مطالب تکمیلی

پیوستهای خود را در صورت وجود میتوانید در ایئو قسمت قرار دهید