

УДК 004.382.2

ДОСЛІДЖЕННЯ ПРОЦЕСОРІВ INTEL ДЛЯ ВИСОКОПРОДУКТИВНИХ ОБЧИСЛЮВАЛЬНИХ СИСТЕМ МЕТОДОМ КЛАСТЕРНОГО АНАЛІЗУ

Литвиненко Д. С., Рябцев В. Г., Люта М. В.

Київський національний університет технологій та дизайну

У цій статті досліджені процесори фірми Intel для високопродуктивних обчислювальних систем методом кластерного аналізу, що дозволяє розбити задану вибірку об'єктів на підмножини так, щоб кожен кластер складався з схожих об'єктів, а об'єкти різних кластерів істотно відрізнялися.

Ключові слова: кластерний аналіз, процесори Intel, високопродуктивні обчислювальні системи

Сучасні засоби багатопроесорної техніки призначені для вирішення задач математичного моделювання, спрямовані на збільшення числа процесорних елементів, що, в свою чергу, опосередковано спричиняє збільшення складності організації комунікацій між цими процесорними елементами. На сьогоднішній день в СНД зареєстровано 13 багатопроесорних систем із загальним числом процесорних елементів більше 1000. У світі налічується 291 система з числом процесорів від 4 до 8 тисяч, а також 147 систем, де кількість процесорів перевищує 8 тисяч [1]. Складності виникають вже при аналізі комунікаційного середовища для тисячі процесорів, а зі збільшенням числа процесорів складності обробки та аналізу ростуть багаторазово.

Постановка завдання

Число процесорних елементів в суперкомп'ютерах постійно зростає, що призводить до зростання складності організації комунікаційного середовища. З метою мінімізації втрат продуктивності від затримок в комунікаційному середовищі і полегшенню розуміння особливостей поведінки комунікаційного середовища застосовуються різні підходи до збору статистичної інформації про передачу повідомлень і подальшому аналізу зібраної інформації. У ході виконання аналізів тестування створюється величезний об'єм даних, який неможливо проаналізувати вручну. Розглядаються задачі стиснення, візуалізації, автоматичного аналізу таких даних, які пропонується вирішувати за допомогою методів кластеризації.

Як і будь-який інший метод, кластерний аналіз має певні недоліки і обмеження: Зокрема, склад і кількість кластерів залежить від обраних критеріїв розбиття. При зведенні вихідного масиву даних до більш компактному увазі можуть виникати певні спотворення, а також можуть губитися індивідуальні риси окремих об'єктів за рахунок заміни їх характеристиками узагальнених значень параметрів кластера. При проведенні класифікації об'єктів ігнорується та дуже часто відсутні в розглянутій сукупності будь-які значення кластерів.

Об'єкт і метод досліджень

В даному дослідженні використовується кластерний аналіз для високопродуктивних обчислювальних систем, який дозволяє проаналізувати та розбити процесори по категоріям, до яких вони відносяться [2-3].

На сьогодні представлено широкий вибір процесорів Intel для виконання різноманітних задач. Всі вони орієнтовані під різні задачі та потреби користувачів, будь то для виконання офісних задач чи для домашнього користування, або ж це ігрова чи мультимедійна станція, а також сервер.

У сучасних процесорах Intel непросто розібратися навіть фахівцеві: випускається безліч різноманітних моделей, а їх назви ніби спеціально покликані заплутати покупця. І якщо про серії Intel Core і Intel Core 2 за майже п'ять років з моменту їх появи написано достатньо, то про чіпах трьох новітніх сімейств Core i3, i5 і i7 практично немає систематизованої інформації, адресованої споживачеві, а не експерту [4].

Кожна нова архітектура процесорів Intel дуже відрізняється та має ряд відмінностей від своїх попередників. І тому постає питання який саме процесор краще та який обрати.

Результати досліджень і їх обговорення

Для вирішення цієї проблеми і використовується кластерний аналіз. А саме за допомогою програми «STATISTICA» [5-7]. Ця програма приймає вхідні дані – характеристики, які потім за допомогою аналізу ми бачимо результат у вигляді кластерів які відображають до якого типу обчислювальних систем відноситься процесор. Спираючись на ці дані користувач може обрати процесор який йому необхідний.

STATISTICA – це система для статистичного аналізу даних, що включає широкий набір аналітичних процедур і методів: більше 100 різних типів графіків, описові та внутрішньо групові статистики, розвідувальний аналіз даних, кореляції,

швидкі основні статистики і блокові статистик, інтерактивний імовірнісний калькулятор, Т-критерії (і інші критерії групових відмінностей), таблиці частот, спряженості, прапорів і заголовків, аналіз багатовимірних відгуків, множинна регресія, непараметричні статистики, загальна модель дисперсійного і коваріаційного аналізу, підгонка розподілів, видобуток даних, нейронні мережі та багато іншого.

Продукти серії STATISTICA засновані на найсучасніших технологіях, повністю відповідають останнім досягненням в області ІТ, дозволяють вирішувати будь-які завдання в області аналізу і обробки даних, ідеально підходять для вирішення практичних завдань в маркетингу, фінансах, страхуванні, економіці, бізнесі, промисловості, медицині і т. д. [7-10]. На рис. 1 наведено вертикальну дендрограму на основі вхідних даних.

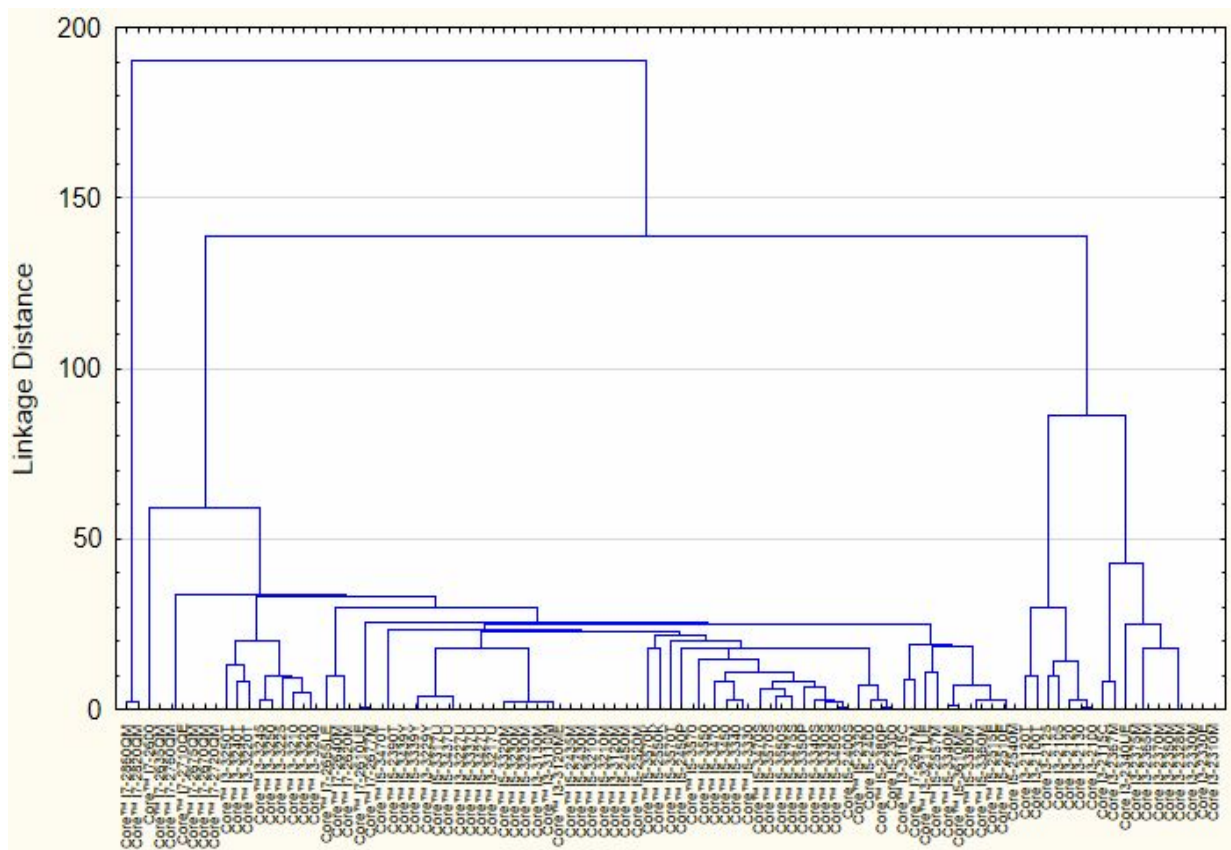


Рис. 1. Вертикальна дендрограма

На рис. 2 приведено, що найбільшим вкладом рішень при кластеризації є «Кількість ядер» процесорів Intel. Виходячи з рівня значимості всі змінні вносять значний вклад в кластеризацію.

Variable	Between SS	df	Within SS	df	F	signif. p
Кол-во ядер	142073,3	4	51,3	94	65084,7	0,000000
Кол-во потовков	137415,0	4	15,0	94	215283,5	0,000000
Тактовая частота GHz	8,9	4	24,0	94	8,7	0,000005
Кэш-память MB	87,0	4	136,0	94	15,0	0,000000
Макс. расч. Мощность W	15339,2	4	35236,2	94	10,2	0,000001
Рекомендуемая цена для покупателей \$	496831,0	4	156884,7	94	74,4	0,000000

Рис. 2. Дисперсійний аналіз, метод К-середніх

Розбиття вхідних даних на кластери виконано методом к-середніх. На рис. 3 показано результати кластеризації. В I кластері знаходяться процесори для вирішення мультимедійних задач, у II кластері – процесори для ігрових станцій, у III кластері – процесори для серверів, станцій по обробці відеоматеріалів та графічних розрахунків та задач підвищеної складності, у IV кластері – для домашнього користування, та для вирішення задач невисокої складності, V – кластері для офісних потреб, де відсутні великі навантаження на роботу процесора.

1-кластер	Members and Distar Cluster co Distance	2-кластер	Members and Distar Cluster cc Distance	3-кластер	Members and Distar Cluster co Distance	4-й кластер	Members and Distar Cluster co Distance	5-й кластер	Members and Distar Cluster co Distance
Core™ i5-2520M	1,06414	Core i5-2540M	11,31231	Core™ i7-2720QM	19,49492	Core™ i5-2380P	14,53210	Core i3-2310M	12,29959
Core™ i5-2430M	1,06305	Core™ i5-2510E	11,31213	Core™ i7-2820QM	58,89301	Core™ i3-3240	18,16524	Core i3-2100	33,63905
Core™ i5-2450M	1,06414	Core™ i5-2515E	11,31213	Core™ i7-2635QM	19,49510	Core™ i3-3225	14,60665	Core i3-2100T	27,84905
Core™ i5-2557M	13,50746	Core™ i7-2600	25,43200	Core™ i7-2670QM	19,49492	Core™ i3-3220	18,16494	Core i3-2120	32,48124
Core™ i5-2435M	1,06305	Core™ i7-2620M	21,39842	Core™ i7-2675QM	19,49492	Core™ i3-3220T	21,47834	Core i3-2120T	31,92616
Core™ i5-2550K	24,09502	Core™ i7-2640M	21,39866	Core™ i7-2710QE	19,49497	Core™ i3-3240T	24,24117	Core i3-2130	33,64038
Core™ i3-3217U	8,24537	Core™ i7-2677M	11,56813	Core™ i7-2760QM	19,49497	Core™ i3-3210	20,15957	Core i3-2330E	12,29932
Core™ i3-3120ME	1,06305	Core™ i7-2637M	6,80767	Core™ i7-2860QM	58,07668	Core™ i3-3250	9,62327	Core i3-2330M	12,29932
Core™ i3-3110M	1,06305	Core™ i7-2655LE	21,63074			Core™ i3-3250T	17,36488	Core i3-2350M	12,29919
Core™ i3-3120M	1,06414	Core™ i7-2610UE	11,57151			Core™ i3-3245	10,74540	Core i3-2370M	12,29919
Core™ i3-3229Y	9,87933	Core™ i3-3217UE	14,84207			Core™ i5-3330	9,13255	Core i3-2340UE	24,11142
Core™ i3-3227U	8,24429	Core™ i3-3115C	3,27963			Core™ i5-3330S	5,59938	Core i3-2105	27,15000
Core™ i3-3130M	1,06680	Core™ i5-3360M	10,09263			Core™ i5-3450	8,14092	Core i3-2125	23,45511
Core™ i5-3320M	1,62079	Core™ i5-3610ME	7,24975			Core™ i5-3450S	6,41069	Core i3-2367M	40,94184
Core™ i5-3427U	8,24537	Core™ i5-3380M	10,09332			Core™ i5-3475S	13,33513	Core i3-2115C	40,35712
Core™ i5-3550	16,90049	Core™ i5-3340M	6,84475			Core™ i5-3470T	13,73060	Core i3-2365M	15,10254
Core™ i5-3550S	13,76447					Core™ i5-3470S	10,88882	Core i3-2328M	12,29932
Core™ i5-3570K	16,95199					Core™ i5-3470	11,98975	Core i3-2375M	15,10154
Core™ i5-3570T	8,38371					Core™ i5-3350P	3,98104		
Core™ i5-3570S	13,76528					Core™ i5-3340	9,13259		
Core™ i5-3570	16,90145					Core™ i5-3340S	7,63064		
Core™ i5-3317U	8,24664								
Core™ i5-3210M	1,06414								
Core™ i5-3210M	1,06414								

Рис. 3. Розбиття процесорів на 5 кластерів

На рис. 4 зображено графік середніх значень для кожного кластера. Процесори які знаходяться в 1-4 кластерах мають більшу кількість ядер, та кількість потоків ніж процесори в п'ятому кластері. Тактова частота та кеш-пам'ять у них співпадає, максимальна розрахункова потужність у них різниться ненабагато. Процесори які відносяться до 4-го кластеру є самими дешевими з представлених, а ті, які відносяться до 3-го кластеру, є самими дорогими.

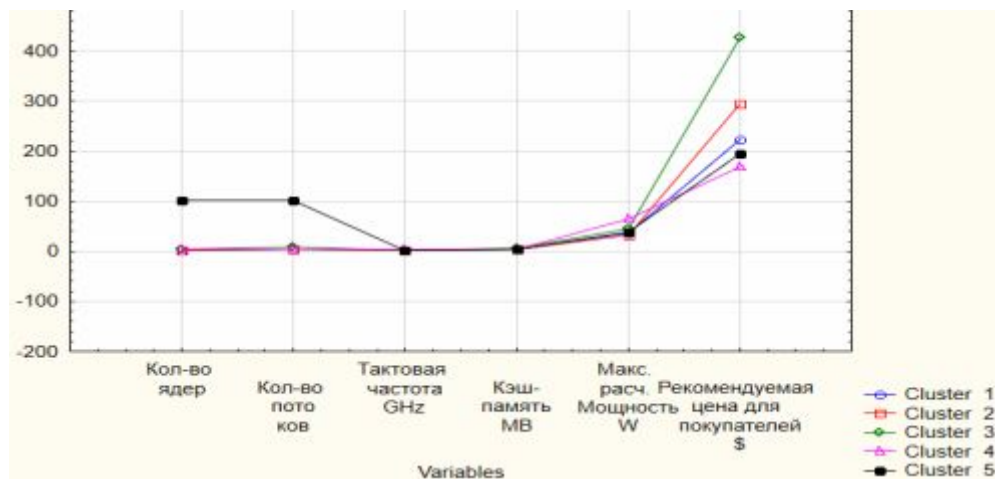


Рис. 4. Графік середніх для кожного кластера

Висновки

У програмі STATISTICA виконано кластерний аналіз процесорів Intel. Він дозволив провести розбиття об'єктів не по одному параметру, а по цілому набору характеристик. Результати роботи мають велике значення в проведенні аналітичних досліджень завдяки можливості перетворити великий обсяг різнобічної інформації в упорядкований, компактний вигляд. Це сприяє підвищенню рівня наочності, зрозумілості та сприйняття результатів аналізу, а також створює підґрунтя для прогнозування. Результати аналізу пропонується використовувати при створенні сучасних засобів багатопроцесорної техніки.

ЛІТЕРАТУРА

1. Брэй Барри. Микропроцессоры Intel: 8086/8088, 80186/80188, 80286, 80386, 80486, Pentium, Pentium Pro Processor, Pentium 4. Архитектура, программирование и интерфейсы. – Петербург : БХВ, 2005. – 1328 с.
2. Воронин А. В. Использование кластерного анализа для выбора локальных стратегий [Электронный ресурс] : Проблемы и перспективы управления

- економікой и маркетингом в организации, 2001. – №1. – Режим доступа : <http://perspectives.utmn.ru/No1/text02.shtml>
3. Мандель И. Д.. Кластерный анализ. – Москва : Финансы и статистика, 1988. – 176 с: ил.
 4. Подробнее о номерах процессоров Intel® [Электронный ресурс] : Номера процессоров – Режим доступа : Intel http://www.intel.com/products/processor_number/rus/
 5. Александров В. В. Анализ данных на ЭВМ. (На примере системы СИТО) / Алексеев А. И., Горский Н. Д. – Москва : Финансы и статистика, 1990. – 192 с.
 6. Хайдуков Д. С. Применение кластерного анализа в государственном управлении. Философия математики: актуальные проблемы. – Москва : МАКС Пресс, 2009. – 287 с.
 7. Дюран Б. Кластерный анализ. – Москва.: Статистика, 1977. – 128 с.
 8. Факторный, дискриминантный и кластерный анализ: Пер с англ. / Дж. О. Ким, Ч. У. Мюллер, У. Р. Клекка и др. Под ред. И. С. Енюкова. – Москва : Финансы и статистика, 1989. – 215 с.
 9. Методы анализа данных: Подход, основанный на методе динамических сгущений: Пер. с фр. / Кол. Авт. Под рук. Э. Дидэ; Под ред. и с предисл. С. А. Айвазяна и В. М. Бухштабера. – Москва : Финансы и статистика, 1985. – 357 с.
 10. Загоруйко Н. Г. Пакет прикладных программ ОТЭКС (для анализа данных). – Москва: Финансы и статистика, 1986. – 160 с.

Литвиненко Д. С., Рябцев В. Г., Лютая М. В.

Исследование процессоров Intel для высокопроизводительных вычислительных систем методом кластерного анализа

В этой статье исследованы процессоры фирмы Intel для высокопроизводительных вычислительных систем методом кластерного анализа, позволяет разбить заданную выборку объектов на подмножества так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались.

Ключевые слова: кластерный анализ, процессоры Intel, высокопроизводительные вычислительные системы

Lytvynenko D. S., Ryabtcev V. G., Lyuta M. V.

The research of Intel processors for high-performance computing systems by cluster analysis

This article investigates processors from Intel for systems by means of cluster analysis. It enables to split the selected sample of objects into subsets so that each cluster consists of similar objects, and objects of different clusters are significantly different.

Keywords: cluster analysis, processors Intel, high-performance computing systems