Цахим сургалтын оюутны сэтгэл ханамжийн кластер шинжилгээ: K-means & divisive hierarchical

Т.Ахыт[1], Л.Пүрэвдолгор[1], Ц.Жавзмаа[1], П.Жаргалбат[1], Ч.Галбадрах[1], Л.Ажнай[1], Ч.Баасандорж[1]

[1]АШУҮИС, Био-Анагаахын сургууль, Физик, Мэдээлэл зүйн тэнхим
Цахим шуудан: akhit@mnums.edu.mn, Утас: 99049542

Түлхүүр үг:

Асуумж судалгаа
Өгөгдлийн тандалт
Задалж-шатлах арга

Товч утга: Их, дээд сургуулийн үйлчилгээний чанарын шалгуур үзүүлэлтийн нэг бол оюутны сэтгэл ханамж юм. Оюутнууд хичээлд оролцох идэвх, сургалтын орчин, хичээлийн үзүүлэн, материаллаг бааз, багшийн ур чадвар, хандлагад хэр сэтгэл ханамжтай байгааг судлахаар зорилоо. АШУҮИС-ийн суурь боловсролын I дамжааны анагаахын физик хичээлийг судалсан нийт 502 оюутнаас гурван бүлэг бүхий нийт 41 асуулттай асуумж судалгаа авлаа. Үүнд I бүлэгт тухайн хичээл хэрхэн үр дүнтэй, сонирхолтой байсан тухай 10 асуулт, II бүлэгт багшийн ур чадвар, хандлагын тухай 19 асуулт, III бүлэгт оюутан өөрөө хичээлдээ хэрхэн оролцсон тухай 12 асуулт бүхий сэтгэл ханамжийн судалгааны 502 бичлэгтэй мэдээллийн сан үүсгэв. Энэхүү мэдээллийн сандаа k-means, задалж, шатлах (Divisive Hierarchical Clustering) аргын R програмын eclust() функцээр кластер шинжилгээ хийлээ. Сэтгэл ханамжийг “маш сайн”, “сайн”, “дунд”, “муу”, “маш муу” гэсэн таван түвшингээр үнэлэв. Судалгаанд k-means аргаар 0.1-0.4, задалж-шатлах аргаар хуваах коэффициент 0.88-0.91 гарлаа. Иймээс задалж-шатлах аргаар хичээлийг үнэлэх гурав, багшийг үнэлэх гурав, оюутны хичээлд оролцоог үнэлэх хоёр кластерт тус тус хуваалаа. Хичээлийн үнэлэх загварын I кластерт дунд үнэлгээ 33%, II кластерт сайн үнэлгээ 61.4%, III кластерт муу үнэлгээ 5% байна. Багшийг үнэлэх загварын II кластерт сайн үнэлгээ 63.9%, I кластерт дунд үнэлгээ 31.8%, III кластерт муу үнэлгээ 4.1% байна. Оюутны хичээлийн идэвхийг үнэлэх загварын I кластерт сайн үнэлгээ 69.9%, II кластерт дунд үнэлгээ 30.07% байна. Шинжилгээнд задалж-шатлах загваруудын хуваах коэффициент нь илүү ач холбогдолтой загвар бүхий кластер болохыг илтгэж байна. Энэхүү хичээлийг үнэлэх загварт 61.4% сайн, багшийн үнэлгээний загварт 69.1% сайн, оюутны идэвх, оролцооны загварт 31.2 маш сайн, 38.8 сайн байна.

Үндэслэл: Боловсрол нь нийгмийн хөгжилд түлхэц үзүүлдэг гол хүч бөгөөд энэ салбарын хөгжилд багагүй хөрөнгө зарцуулсаар байна. Зарцуулж буй хөрөнгийн үр ашгийг нэмэгдүүлэхэд чанарын асуудал тавигдаж, боловсролыг хэрхэн чанартай, илүү үр ашигтай болгох, түүний судалгааг хөгжүүлэх талаар боловсролын бүх түвшинд чанарын үнэлгээ хийсээр байна. Их сургуулийн үйлчилгээний чанарын шалгуур үзүүлэлтийг оюутны сэтгэл ханамжаар үнэлдэг юм.

Зорилго: Анагаахын суурь боловсролын дамжааны анагаахын физикийн хичээлийг судалж буй оюутнуудаас сургалт, багшийн хандлага ур чадвар, материаллаг баазын хангамжид хэр сэтгэл ханамжтай байгааг судлах зорилгоор дараах зорилтуудыг дэвшүүлж ажилласан. Үүнд:

1. Кластерын задалж шатлах, k-means аргыг судлах
2. Анагаахын суурь боловсролын дамжааны анагаахын физикийн хичээлийг судалж буй оюутнуудаас сургалт, багшийн хандлага ур чадвар, сургалтын орчинд сэтгэл ханамжтайг тодорхойлох

Өгөгдлийн тандалт: Өгөгдлийн тандалт нь их хэмжээний өгөгдлөөс далд зүй тогтол, уялдаа холбоог нээн илрүүлэх процесс бөгөөд өгөгдлийн сангийн систем, математик, статистик болон загварчлах аргуудаар мэдлэг олж авах аргачлал юм[1-3]. Өгөгдлийн тандалтад цугларсан их хэмжээний өгөгдлүүдийн хоорондын ерөнхий шинжийг тодорхойлох, өгөгдлүүдийн холбоо, хамааралд дүн шинжилгээ хийх, өгөгдлүүдийг нийтлэг шинжүүдээр нь бүлэглэх, тодорхой үзүүлэлтүүдээр урьдчилан таамаглах, хэт ялгаатай өгөгдлүүдийг ялгах гэх мэт тооцоолол хийхийн тулд ангилал (Classification) болон бүлэглэлийн (clustering) аргуудыг ашигладаг.

Кластер шинжилгээ: Бүлэг дэх объектууд өөр хоорондоо адил, бусад бүлгийн объектоос өөр байхаар багцын объектуудыг хуваж, байрлуулахыг кластер шинжилгээ (cluster analysis) гэнэ. Кластер шинжилгээнд хуваалтын (Partitioning) ба шатлах (Hierarchical) аргууд байна[4,5].

Хуваалтын k-means арга: n объектыг агуулсан D багцыг $C_1 \ldots C_k$ гэх мэт k кластерт $C_i \subset D$ ба $C_i \cap C_j = 0$ нөхцөлтөйгөөр хуваж байрлуулна. Хуваж байрлуулахдаа кластер бүрийн хувьд тухайн объектоос дундаж объект хүртэлх зайг квадрат алдааны нийлбэрээр тооцно[6-8].

Өгөгдлийн сангийн объектуудыг дооройс дээшээ нэгтгэх (bottom-up), дээрээс доошоо (top-down) хуваах, задлах загвараас хамаарч шаталсан задаргаа үүсгэдэг. Шаталсан кластер (hierarchical clustering) нь нэгтгэх (agglomerative), задлах (divisive) аргуудтай.

Нэгтгэж, шатлан кластерчлах (agglomerative hierarchical clustering): Өгөгдлийн сангийн обьектуудын атрибут бүрээр кластер үүсгэнэ. Эдгээр кластеруудыг алхам бүрд дооройс дээшээ нэгтгэж, хамгийн сүүлийн алхамд нэг багц кластер үүсгэдэг. Энэ нь ихэвчлэн объект бүр өөрийн кластер үүсгэж эхэлдэг бөгөөд кластеруудыг давтамжтайгаар илүү том кластеруудад нэгтгэдэг бөгөөд бүх объектуудыг тодорхой нөхцөл дуусах хүртэл нэг кластерт нэгтгэнэ[4,9-11].

Задалж, шатлан кластерчлах (divisive hierarchical clustering): Бүх обьектуудыг шатлалын эх гэх нэг кластерт байрлуулна. Обьектуудын хоорондын хамгийн бага, хамгийн их, дундаж зайгаар ижил төстэй шинжээр нь жижиг кластеруудад дээрээс доошоо зарчмаар хуваана. Эхлээд шатлалын орой болох ганц кластерт бүх объектуудыг байрлуулна. Орой кластераа жижиг дэд кластеруудад хуваана. Хуваагдсан кластерууд дахин ганц жижиг бүлгүүдэд өөрөө өөрийгөө задлана. Хуваах процесс нь хамгийн бага түвшний кластер бүрийн объектууд хоорондоо хангалттай төстэй, уялдаа, холбоотой болох хүртэл үргэлжилнэ[4,10,11]. R программын eclust функцээр задалж, шатлах кластер болон k-means аргаар бүлгүүдэд хуваадаг[4,12].

Кластерын үнэлгээний дүрсийн коэффициент: Дотоод аргын дүрсийн коэффициентоор (silhouette coefficient) үүссэн кластерыг үнэлнэ[4]. Дүрслэлийн коэффициентыг (1) томьёогоор тооцов.

$$s(i) = \frac{(b(i) - a(i))}{max\{a(i), b(i)\}}$$

Дээрх (1) тэгшитгэлд a(i)- i объектыг агуулж байгаа кластерын объект хоорондын дундаж зай (кластер доторх зай), b(i) - i объектыг агуулж кластерын объект ба өөр кластерын объект хоорондын хамгийн бага зай болно. Дүрсийн коэффициент нь -1<s(i)<1 хооронд байх бөгөөд хүчтэй, зохистой, сул бүтэц байгаа эсэхийг илэрхийлнэ.

Хуваалтын коэффициент: Хуваалтын коэффициентоор (divisive coefficient) олдсон шаталсан кластерын бүтцийг хэмждэг. 0.5-с дээш үед хүчтэй сайн бүтэц олдсоныг илэрхийлэв.

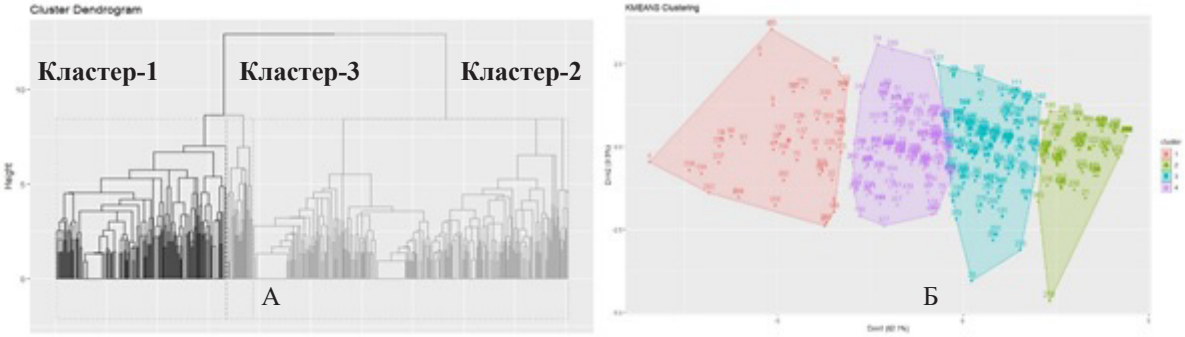
$$Dc = 1 - \mathrm{d(i)}$$

Дээрх (2) тэгшитгэлд i объект бүрийн хувьд d(i) нь харьяалагдах сүүлийн кластерын диаметрийг том багцалсан кластерын диаметрт хуваж тодорхойлов.

Кластер шинжилгээний үр дүн: Анагаахын суурь боловсролын I дамжааны 502 оюутнаас анагаахын физикийн хичээлээр 10 атрибуттай хичээлийг үнэлэх, 19 атрибуттай багшийн оролцоог үнэлэх, 21 атрибуттай оюутныг үнэлэх мэдээллийн сангууд үүсгэн, R программын кластерын задалж шатлах, k-means аргуудаар бүлгүүдэд ижил төстэй шинжээр хуваж, загваруудыг тооцоолов.

Загвар тус бүрд кластерын тоо, дүрслэлийн коэффициент, хуваалтын коэффициент, объектуудын тоо, сэтгэл ханамжийн түвшин зэрэг үр дүн харгалзана. Үүссэн кластеруудын дүрсийн коэффициент s(i) нь 0.1-0.4 хооронд дундаж, шаталсан аргын хуваалтын коэффициент, Dc нь 0.88-0.91 хооронд хүчтэй бүтэц олдсоныг илтгэж байна.

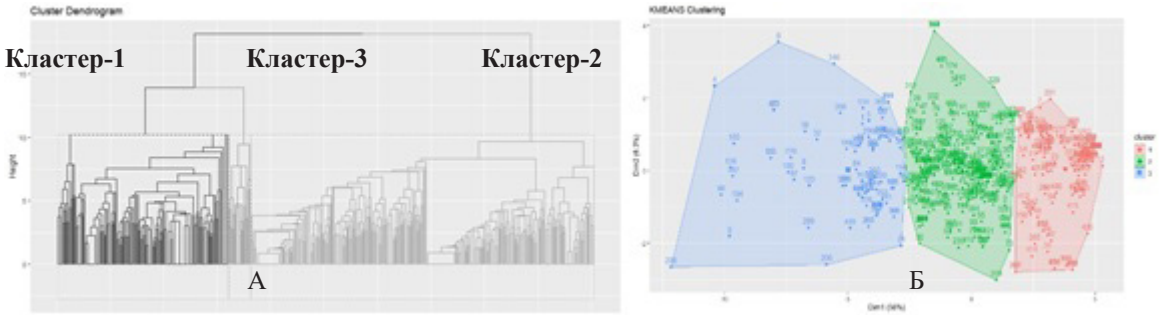
Хичээлийг үнэлэх загвар: Хичээлийн үнэлэх санг гурван кластерт задалж шатлах аргаар хуваасныг Зураг 1А-д дүрслэв. Сэтгэл ханамжийн түвшин дунд 167 бичлэгтэй I кластер, сэтгэл ханамжийн түвшин сайн 309 объекттой II кластер, сэтгэл ханамжийн түвшин муу 26 объекттой III кластерыг илэрхийлнэ.



Зураг 1. Шаталсан (А), k-means (Б) аргаар тооцоолсон хичээлийг үнэлэх загвар

Сэтгэл ханамжийн түвшин муу 50 бичлэгтэй I, сэтгэл ханамжийн түвшин маш сайн 102 объекттой II, сэтгэл ханамжийн түвшин сайн 190 объекттой III, сэтгэл ханамжийн түвшин дунд 60 объекттой IV кластерыг илэрхийлэв (Зураг 1Б).

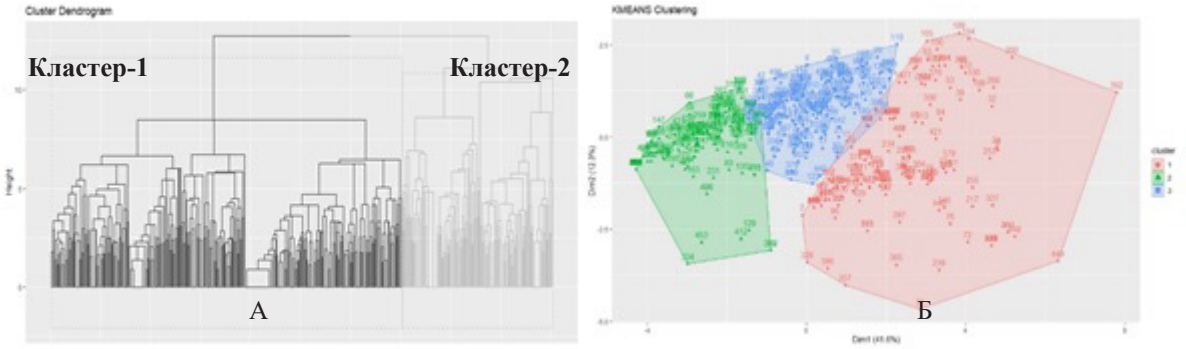
Багшийн оролцоог үнэлэх загвар; Шаталсан аргаар гурван кластерт хуваагдсан багшийн оролцоог үнэлэх загварыг Зураг 2А-д харуулав. Сэтгэл ханамжийн түвшин сайн 321 бичлэгтэй II кластерыг дүрсэлж байна. Сэтгэл ханамжийн түвшин дунд 160 объекттой I кластер, сэтгэл ханамжийн түвшин муу 21 объекттой III кластерыг илэрхийлэв.



Зураг 2. Шаталсан (А), k-means (Б) аргаар тооцоолсон багшийн оролцоог үнэлэх загвар

Зураг 2Б-д k-means аргаар хуваасан үр дүнг үзүүлэв. Сэтгэл ханамжийн түвшин маш сайн 147 бичлэгтэй I, сэтгэл ханамжийн түвшин сайн 264 объекттой II, сэтгэл ханамжийн түвшин дунд 91 объекттой III кластерыг илэрхийлнэ.

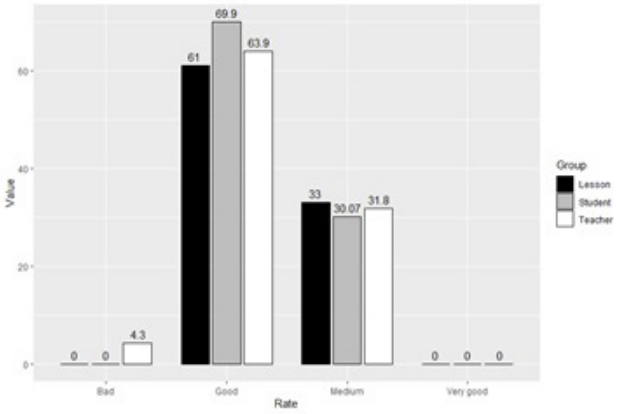
Оюутныг үнэлэх загвар: Шаталсан аргаар тооцсон хоёр кластерт хуваагдсан оюутныг үнэлэх загварыг Зураг 3А-д харуулав. Сэтгэл ханамжийн түвшин сайн 311 бичлэгтэй I кластерыг, сэтгэл ханамжийн түвшин дунд 151 объекттой II кластерыг илэрхийлж байна.



Зураг 3. Шаталсан (А), k-means (Б) аргаар тооцоолсон оюутныг үнэлэх загвар

Сэтгэл ханамжийн түвшин дунд 159 бичлэгтэй I, сэтгэл ханамжийн түвшин маш сайн 149 объекттой II, сэтгэл ханамжийн түвшин сайн 194 объекттой III кластерыг дүрсэлж байна (Зураг 3Б).

Шаталсан кластерын аргын хуваах коэффициентоор хүчтэй бүтэц олдсон тул хичээл, багшийн оролцоо, оюутныг үнэлэх загваруудын үр дүнг нэгтгэв (Зураг 7). Анагаахын физик хичээлийн сэтгэл ханамжийн түвшин 5% муу, 61% сайн, 33% дунд, багшийн оролцоо 3% муу, 69% сайн, 27% дунд, оюутнуудын оролцоо, идэвх 39% сайн, 29% дунд, 31% маш сайн гэж үнэлсэн байна.



Зураг 4. Шаталсан аргаар тооцсон хичээл, багшийн оролцоо, оюутныг үнэлэх загварын үр дүн

Дүгнэлт:

1. Анагаахын суурь боловсролын дамжааны анагаахын физик хичээлийг судалж буй оюутнуудын хичээл, багшийн оролцоо, оюутны сэтгэл ханамж зэргийг олон улсад өргөн хэрэглэгддэг кластерын задалж шатлах, k-means аргаар судлахад задалж-шатлах аргын хуваах коэффициент 0.88-0.91 хооронд байгаа нь ач холбогдолтой бүтэц олдсоныг илтгэж байна. Иймд задалж шатлах аргыг судалгаа шинжилгээ ашиглах боломжтойг харуулж байна.
2. Хичээлийг үнэлэх загвараар физикийн хичээлд 61.4% сайн, багшийн үнэлгээний загвараар багшид 69.1% сайн, оюутны идэвх, оролцооны загварт оюутнууд өөрсдийгөө 31.2 маш сайн, 38.8 сайн үнэлжээ.

Ном зүй:

1. *Сарангэрэл Д. Өгөгдлийн тандалт болон үүлэн тооцооллын зарим арга зүйг эрүүл мэндийн салбарт нэвтрүүлэх судалгаа. УБ, ШУТИС; 2013.*
2. *Hand DJ. Principles of data mining. Drug safety. 2007;30(7):621-622.*
3. *H Witten I. Data mining. 2016.*
4. *Han J, Pei J, Kamber M. Data mining: concepts and techniques. Elsevier; 2011.*
5. *Tan P-N, Steinbach M, Kumar V. Introduction to Data Mining. 2006.04.25.*
6. *Hartigan JA, Wong MA. Algorithm AS 136: A k-means clustering algorithm. Journal of the royal statistical society series c (applied statistics). 1979;28(1):100-108.*
7. *Likas A, Vlassis N, Verbeek JJ. The global k-means clustering algorithm. Pattern recognition. 2003;36(2):451-461.*
8. *science Td. K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks.*
9. *Sasirekha K, Baby P. Agglomerative hierarchical clustering algorithm-a. International Journal of Scientific and Research Publications. 2013;83:83.*
10. *Murtagh F, Contreras P. Algorithms for hierarchical clustering: an overview. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2012;2(1):86-97.*
11. *Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. Vol 344: John Wiley & Sons; 2009.*
12. *RDocumentation. Visual Enhancement Of Clustering Analysis. https://www.rdocumentation.org/packages/factoextra/versions/1.0.5/topics/eclust.*

Student Satisfaction of E- Training based Cluster Analysis: K-means & divisive hierarchical

Akhiit T[1], Purevdolgor L[1], Javzmaa Ts[1], Jargalbat P[1], Galbadrakh Ch[1], Ajnai L[1], Baasandorj Ch[1]

[1]MNUMS
Email: akhit@mnums.edu.mn, Tel: 99049542

Student satisfaction is considered as one of the university education service quality indicators.
We aimed to figure out how satisfied the students were with their attendance, learning environment, visuals materials, facilities, teacher skills and attitudes.
A questionnaire with a total 41 items grouped in three were taken from 502 first-year, pre-medical students at the Mongolian National University of Medical Sciences. In the first group of questionnaire consisted of 10 items about how effective and interesting the lesson was; the second group 19 items were regarding to the teacher skills and attitudes, and the last 12 questions were asked how students participated in the lesson. Divisive Hierarchical Clustering method, the eclust function of the R program was employed. The satisfaction was assessed on five levels: “excellent”, “good”, “average”, “bad”, and “very bad”.
In the study, the k-means method was 0.1-0.4 and the divisive hierarchical factor was 0.88-0.91. Therefore, we formed three clusters for classroom assessment (I), three for teacher evaluation (II), and two for student participation (III). The average grade is 33% in cluster I, 61.4% is good in cluster II, and 5% is bad in cluster III. The teacher evaluation model has a good grade of 63.9% in cluster II, an average grade of 31.8% in cluster I, and a poor grade of 4.1% in cluster III. The good grade in cluster I was 69.9% and the average grade was 30.07% in cluster II.
The analytical study demonstrated that the divisive coefficient is more significant cluster model. The 61.4% was good in the subject evaluation model, 69.1% - good in the teacher evaluation model, and the student active and participation model was very good (31.2) and good (38.8).

Key word: Survey, data mining, cluster analysis