

Хадан Дээрх Тамганы Дүрсийг Адууны Тамганы Дүрстэй Машин Сургалтын Аргуудаар Харьцуулах

Пэрэнлэйлхүндэв Гантуяа*, Батсуурь Сувдаа*, Дамдинсүрэн Цэвээндорж**

*Монгол Улсын Их Сургууль, Мэдээлэл, Компьютерийн Ухааны Тэнхим

**Шинжлэх Ухааны Академи, Түүх Археологийн Хүрээлэн

И-мэйл: gantuya@seas.num.edu.mn, suvdaa@seas.num.edu.mn, dtseveen@yahoo.com

Хураангуй: Нүүдлийн соёлын олон зууны түүхэн уламжлалтай манай орны нутагт орших археологийн олдвор болох хад, чулуун дээр дүрслэн үлдээсэн элдэв тамга нь тухайн газарт амьдарч буй малчдын адуунд тамга хэлбэрээр уламжлагдан ирсэн дүрсийн хамаарлыг бидний хийсэн судалгаанаас харж болно. Энэ ажлаар бид адууны 244 тамгыг Хэнтий аймгийн нутагт орших Рашаан хад хэмээх газарт байх том чулуунд сийлж үлдээсэн 260 гаруй тамганы дүрстэй харьцуулах ажлыг хийсэн. Ингэснээр монгол малын тамганы угшил тухайн газар нутагт байх археологийн олдвор, хадны зургаас үүдэлтэй болох талаар санааг дэвшүүлэн хамаарлыг судалсан болно. Харьцуулах үйл ажиллагааг машин сургалтын арга KNN, NaiveBayes, Decision Table, random Forest зэрэг үндсэн алгоритмуудаар таниулан туршсан. Эндээс хамгийн өндөр үр дүнгээр нь хадны дүрс адууны тамганы дүрсийн хамаарлыг тодорхойлсон болно.

Түлхүүр үгс: Адууны тамга, хадны дүр, танилт, KNN, NaiveBayes, Decision Table, random Forest

I. ОРШИЛ

Компьютерийн ухаан, тэр дундаа машин сургалтын аргууд нь өнөө цагт олон салбарт хэрэглэгдэх болсон билээ. Бидний энэ удаагийн ажил нь урлаг судлал болон компьютерийн ухааны салбар дундын шинжтэй. Урлаг судлалын талаас нь авч үзвэл өнөөгийн монгол үндэстний адууны тамганаас ч эртний монголчуудын соёлын шинж улбааг олж харж болох юм. Нэг үгээр хэлбэл Монголчуудын тамгаар дамжсан соёлын улбаа тасралтгүй тов тодоор үлджээ. Хэнтий аймгийн Батширээт сумын нутаг Рашаан хаднаа сийлбэрлэгдэн үлдсэн асар баян тамга дүрсээс жишээ болгон авч туршилт хийсэн. Дэлхийн улс түмэн, овгийн байгуулалтай байх үеэсээ эхлэн нэг овог нөгөөгөөс ялгах тэмдэг, билэгдэл хэрэглэх болжээ. Тэр тамга тэмдгээ идээшиж нутагшсан газрынхаа сүр бараатай хад чулуу болон өөрсдийн үйлдсэн сав суулга, хөдөлмөрийн багаж, бусад эд өлгөсийн зүйлд дурсгал болгон сийлж үлдээдэг заншилтай байв. Тийм тамга тэмдэг одоо манай орны аль ч нутгаас элбэг олддог юм.

Монгол орны хэмжээнд Хэнтийн Рашаан хад, Говь-алтайн Цагаан гол, Архангайн Тайхар чулуу зэрэг хад цохионд оршиж байсан эртний овгуудын тамгын цуглуулга, баялаг дурсгал үлдснийг судлаачид олж илрүүлсэн байдаг.[1] Тэдгээрийг эрдэмтэд эртний овог аймгийн соёл иргэншлийн хөгжлийг тодорхойлох хэрэглэгдэхүүн гэж үздэг байна. Эдний дотроос хамгийн баян цуглуулгатай нь 260 шахам тамгатай

Рашаан хад юм. Рашаан хад одоогийн Хэнтий аймгийн Батширээт сумын нутаг дэлхийн газарзүйн 110-р уртраг, 48-р өргөрөгт оршино. Хадан самбарын урт нь 6м, өргөн нь 2-3м, зузаан нь 60-70 см буюу нийт зураг дүрс бүхий толио нь 25 хавтгай дөрвөлжин м орчим талбай эзэлсэн байна. Олон дүрс тэмдгийг нэгдсэн тогтолцоотойгоор байрлуулж сийлээгүйн дээр уг чулуу нь бүдүүн ширхэгтэй боржин тул өнгөн тал нь он цагийн нөлөөнд их л холтрон гэмтсэн байна. Монголд археологийн /овгийн тамганы/ дурсгаланг хад чулуунд үлдээгсэдтэй эрт урьдны элэр баларын угшил уламжлал бүхий болох нь ерөнхийдөө мэдэгдсэн гэж болно.[1]

Иймээс бид өнөөгийн машин сургалтын аргаар хадны дүрс болон адууны тамганы хамаарлыг тодорхойлох зорилгоор түгээмэл хэрэглэгддэг аргуудыг сонгон харьцуулсан туршилт хийлээ. Үүнтэй адилаар гадаадын судлаачдын судалгааг авч үзэхэд брэндийн тэмдэг, замын тэмдэглэгээ, үхэр малын тамга зэргийг машин сургалтын аргаар таниулсан судалгаанууд нэлээд байдаг. Жишээ нь Замын тэмдэг илрүүлэх таних [5], дүрс боловсруулалт ашиглан шуудангийн марк таних ялгах[8], Зурган дээр байгаа огноо илрүүлэх таних [9] мөн дүрсийн ирмэгт суурилсан чиглүүлэгчгүй ангилагчаар зургаас текстийн хэсгийг илрүүлэх [10] зэрэг байна. Энэ ажлын 2-р бүлэгт Хадны дүрс, адууны тамганы талаар, 3-р бүлэгт танилт хийх үе шат, танилтанд ашигласан аргуудын ерөнхий шинжийн талаар, 4-р бүлгээр харьцуулсан үр дүнгийн талаар тайлбарлан бичсэн болно.

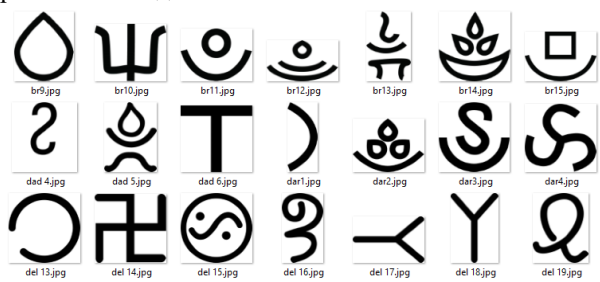
II. ХАДНЫ ДҮРС, АДУУНЫ ТАМГАНЫ ХАМААРАЛ

Бид энэ удаагийн судалгаанд адууны 244 тамганы дүрсийг хадны 260 дүрстэй хамааралтай эсэхийг шалгахар бэлтгэсэн. Түүвэр тус бүрийн зарим жишээг Зураг 1, 2-т дүрслэн үзүүлэв.



Зураг 1. Рашаан хадны тамганы дүрс

Зураг 1-т рашаан хадны тамганы дүрсийг сканердсан зурагнаас сэргээн 10 бүлэгт хуваан ангилж тус бүрд нь зураг болгож хадгалсан.



Зураг 2. Адууны тамга

Зураг 2-т 244 адууны тамга тус бүрийг зураг болгон 10 бүлэгт хуваан, танилтанд бэлтгэж тестийн зураг болгон ангилсан.

III. СИСТЕМИЙН ЕРӨНХИЙ БҮТЭЦ



Зураг 3. Системийн бүтэц

Системийн бүтцийг аваад үзвэл зураг оруулах буюу эхний хэсэгт номын хуудаснаас сканнердсан зураг оруулаад дүрсийг нэг бүрчлэн салгаж *.jpg өргөтгөлтэй хадгална.

Уг зургаас тамганы дүрс нэг бүрийг салган авч тус бүрд зураг болгон хадгалсан. Сургах дүрсүүдийг түүвэрлэн сан үүсгэж, гол компонентын анализ буюу PCA(Principal Component Analysis) [8,9,10] алгоритмыг ашиглан пиксэлийн утга агуулсан 2500 (50x50 хэмжээтэй зургийн) хэмжээст вектороос хамгийн чухал буюу хувийн утга өндөртэй 28 хувийн векторыг сонгон хувиргалт хийн онцлог шинжийг нь гарган авсан. Рашаан хадны 260 дүрсийг k-ойрын хөрш (K-Nearest Neighbors), НэивБэйс(Naive Bayes), Шийдвэрийн мод (Decision Tree), Санамсаргүй ой

(Random Forest), Нейрон сүлжээ (Neural Network) аргуудаар сургалт хийж, сургасан сантайгаа адууны 244 тамгатай харьцуулалт хийсэн.

A. Танилтын аргууд

Бид энэ ажлын хүрээнд энгийн бөгөөд түгээмэл хэрэглэгддэг

1) Хамгийн ойрын хөрш (K-Nearest Neighbors)

Энэ нь хамгийн энгийн бөгөөд боломжийн үр дүн үзүүлдэг ангиллын аргын нэг юм. Сургалтын процесс нь зөвхөн өгөгдлүүдийг хадгалан, танилтын процесст уг хадгалсан өгөгдөл нэг бүртэй оролтын өгөгдлийг харьцуулан хамгийн ойрын буюу хамгийн бага зайтай байгаа k ширхэг өгөгдлийн ихэнх нь аль ангилал байгаагаар шийд гаргадаг. Энэ арга нь энгийн хэдий ч сүүлийн үеийн үр дүнтэй (state-of-the-art) аргуудтай байнга харьцуулагдаж байдаг арга юм. Сургалтын өгөгдөл олон байх тусам, мөн онцлог шинжийн тоо их байх тусам танилтын хугацаа өсдөг нэг сул талтай. [4], [6], [7].

2) NaiveBayes

Энэ арга статистик аргуудын төрөлд ордог ба болж өнгөрсөн процессуудын шинжилгээнд үндэслэн болох процессийг таамаглан танилт хийдэг. Статистик аргууд нь өгөгдлийн түгэлтийн нормал гэж таамаг дэвшүүлэн дундаж олох, дундажаас хазайх хазайлт олох зэргээр нэг ангиллын өгөгдлүүдийн ерөнхий шинжүүд, бусад ангиллаас ялгарах шинж зэргийг голлон тооцож, нөхцөл магадлалын түгэлт, жин зэргээр Бэйсийн онолд үндэслэн танилт хийдэг нэгэн суурь арга юм. Хэрэв өгөгдлийн түгэлт нормал биш байх тохиолдолд алдах эрсдэлтэйгээрээ сул талтай. [11-13]

3) Decision tree

Шийдвэрийн мод нь танилт хийхийн тулд онцлог шинжийн утгуудаар модны зангилаа болгон авч нөхцөл шалган уг нөхцлийн төлвөөс хамааран дараагийн онцлог шинж буюу зангилаанд шилжих замаар бүтэцлэгдсэн байдаг. Танилт хийхдээ модны үндсээс эхлэн навч хүртэл шалгаж буй нөхцлүүдийн биелж байгаа тохиолдолд тодорхой нэг навчинд хүрэх буюу уг навчийг юу гэж ангилсан байгаа тэр ангилал болно. Шийдвэрийн модыг логикийн үйлдлээр бол AND-үүдийн OR-үүд буюу нэг ангилал нь олон навчинд байх боломжтой буюу үндсээс навч хүрэх нэг замын дагуух бүх нөхцлүүд биелсэн байх ёстой гэсэн үг. Шийдвэрийн мод ашиглах үед тулгардаг нэг асуудал нь өгөгдлийн онцлог шинжийн утга дискрет биш байвал асуудал үүсэх юм. Тийм тохиолдолд их буюу тэнцүү, тэдээс бага гэх мэт нөхцөл шалгаж шийдвэрлэдэг. Оновчтой модыг бүтэцлэх нь модны хэмжээ бага жигд тэгш байх юм, энэ тооцооллын хувьд давуу талтай. Үүнийг хэрхэн шийдэхээс хамааран олон төрлийн мод үүсдэг ба энтропи тооцох шаардлагатай байдаг. [14, 15]

4) Санамсаргүй ой (Random Forest)

Санамсаргүй ой арга нь сүүлийн үеийн үр дүн өндөртэй аргуудын нэг бөгөөд ой гэдэг нь олон моднуудаас тогтоно гэсэн утгатай арга юм. Анх 1995 оноос санал болгон гарч ирсэн ба санамсаргүй сонголтоор олон онцлог шинжүүдийн хосмоор шийдвэрийн моднууд үүсгэн, олон боломжуудаас хамгийн оновчтойг сонгох замаар үүсгэнэ. Мөн ойрын хөршүүдтэй ч хослуулан үүсгэх тохиолдол байдаг. Энэ

аргыг танилтын олон ажилд хэрэглэсэн судалгаа их байдаг. [16]

5) *Нейрон сүлжээ (Neural Network)*

Нейрон сүлжээ нь шугаман ангилагчийн өргөтгөсөн загвар гэж хэлж болно. Шугаман ангилагч нь гүдгэр түгэлттэй өгөгдлийн олонлогуудыг шулуунаар ангилах боломжтой буюу математикаар шугаман тэгшитгэлийн хариу боддог шиг сургалтын процесс нь явагддаг. Гэтэл өгөгдлийн түгэлт гүдгэр байх тохиолдол тийм ч их биш учраас шугаман биш байдлаар ангилах нэг хэлбэр нь олон нууц давхарга үүсгээд оролтын утгуудыг тодорхой жин буюу коэффициентээр үржиж нэмээд эхний нууц давхаргын нэг зангилаан дээр хүрнэ, мөн тэр утгыг өөр жингээр үржиж нэмээд дараагийн давхаргын нэг зангилаанд гэх мэт явсаар гаралтын давхаргын зангилаанд хүрч босгочлогдон хариу гардаг. Сургалтын процесс нь олон төрлийн математикийн аргууд ашигладаг ба өнөө үед нейрон сүлжээний маш олон хувилбар гарч, судалгааны олон салбарт үр дүнгээ өгч байгаа. Сургах хугацаа нь удаан ч, таних хугацаа нь хурдан тул онлайн процессуудын их үр дүнтэй хэрэглэгдэж байна. [17] Туршилтын үр дүнг дараах бүлэгт дэлгэрэнгүй тайлбарласан болно.

III. ТУРШИЛТ

Туршилт 1-д Рашаан хадны 260 тамганы дүрсийн 66 хувийг сургаж, 34 хувийг таниулсан үр дүнг, Туршилт 2-д Рашаан хадны 260 тамганы дүрсийг сургаж, Хэнтий аймгийн адууны 244 тамганы дүрсийг РСА-н 28 утган дээр, NaiveBayes, Neural Network (Logistic), DecisionTree, RandomForest, KNN-н аргуудаар харьцуулсан. Үр дүнг нь хүснэгт 1 харуулав.

Хүснэгт 1. РСА-ийн 28 утган дээр танилт хийсэн үр дүн

Туршилт 1		Танилтын аргууд		Зөв ангилсан (Хувь)
		1.	2.	
Рашаан хадны тамга	1.	Naïve Bayes		53 %
	2.	1NN(KNN)		57 %
	3.	Neural Network (Logistic)		41%
	4.	Decision tree(J48)		47%
	5.	Random Forest		53%
Туршилт 2				
Адууны тамга	1.	Naïve Bayes		25%
	2.	Decision tree(J48)		20%
	3.	1NN(KNN)		15%
	4.	Neural Network (Logistic)		24%
	5.	Random Forest		31%

РСА-ийн 28-ийн утган дээр хэдэн хувьтай таньж байгааг хүснэгтээс харьж болно. Эндээс үзвэл хадны дүрсийн хувьд KNN-ийн утга хамгийн их буюу 57%-тай, хадны дүрсийг адууны тамгатай харьцуулахад Санамсаргүй ой арга нь хамгийн өндөр буюу 31%-тай гарч байгаа нь манай судалгааны хувьд үр дүнтэй гэж үзэж байна. Бидний өгөгдлийн хувьд машин сургалтын дээрх аргуудын ойрын хөрш, санамсаргүй ой аргууд нь илүү өндөр танилттай байна.

ДҮГНЭЛТ

Бидний хийсэн судалгаагаар нийт тамганы 30-аас дээш хувьтай таньж байгаа нь малчдын адуундаа дарсан тамга нь хаднаас нилээн хамааралтай гэж дүгнэж болох юм.

Эндээс монголчууд малын тамгаа сонгохдоо олон талын судалгаанд үндэслэн тэр дундаа овгийн тамгийн сэргээн тухайн нутаг оронд байгаа нэр сүртэй хад, чулуун дээр дарсан тамга тэмдгийг баяжуулан сонгодог ард түмэн гэдэг нь харагдаж байгаа юм. Монголчуудын соёл уламжлалыг судлах шинжлэх ухаанд бусад салбарын тэр дундаа компьютерийн ухааны машин сургалтын аргыг ашиглах нь үр дүнтэй юм гэж дүгнэж болно. Цаашид дүрс танилтын бусад аргуудыг судлан бэлчээрт байгаа малын бодит зурагнаас тамгатай хэсгийг ялган авч дүрсээр нь хэний адуу, ямар утга бүхий тамга болох зэргийг мэдэх боломжтой юм.

НОМ ЗҮЙ

[1] Kh.Perlee, “Бүтээлийн чуулган”, Улаанбаатар ISBN: 976-99962-3-125-4, 2012

[2] Ts.Battulga, “Монголын руни бичгийн бага дурсгалууд”, 2005

[3] S.Dulam, “Билэгдэл зүй”, 2010

[4] H.Ebrahimpour and A. Kouzani, “Face Recognition using bagging KNN”, Int’l Conf. on CVPR, IEEE, pp. 209-216, 1996.

[5] H. Fleyeh, “Traffic and Road Sign Recognition”, doctoral thesis, Napier University, 2008

[6] K.Patrik, Martina Z, Robert H, Roman J, Miroslav B and Jan H, “A Novel Approach to Face

[7] Ž.Slovakia Recognition using Image Segmentation Based on SPCA-KNN Method“, University of Žilina, Univerzita 1, 010 26, 2013

[8] K. H.Mohammad., Abu A. S. R., Rajibul A., and Dr.William P., “Automatic Face Recognition System using P-tree and K-Nearest Neighbor Classifier”, 2004

[9] G.LiJune, “Postage stamp recognition using image processing“, Faculty of Engineering and Sustainable Development, 2011

[10] X.Chen and H. Zhang “Photo time-stamp detection and recognition” 7th International Conference on Document Analysis and Recognition (ICDAR 2003), 2-Volume Set, Edinburgh, Scotland, UK, pp.319-322, 2003.

[11] Harry Zhang, "The Optimality of Naive Bayes," in 17th FLAIRS Conference, 2004.

[12] Wray Lindsay Buntine. A Theory of Learning Classification Rules. PhD thesis, University of Technology, Sydney, 1990.

[13] W. J. Frawley, G. Piatetsky-Shapiro, and C. J. Matheus. Knowledge discovery in databases: An overview. Ai Magazine, 13:57–70, 1992.

[14] Aha, D., Kibler, D. & Albert, M. (1991). Instance-based learning algorithms. Machine Learning, 6(1), 37–66.

[15] Ryuichi Oka. Spotting method for classification of real world data. The Computer Journal, 41(8):559–565, 1998

[16] S. Sheik Mohammed Ali, Niranjana Joshi, Bobby George, and Lelitha Vanajakshi. Application of random forest algorithm to classify vehicles detected by a multiple inductive loop system. In 15th International IEEE Conference on Intelligent Transportation Systems, pages 491495, September 2012.

[17] A Global-Local Learning Algorithm, Proceedings from the World Congress on Neural Networks, Vol 1, pp. 587-590, Washington 95. J. Depenau.