

УДК 81'322::811.222.8::519.25

ББК 81.1::Ш152.131.2-923::В172

МУАЙЯНКУНИИ МУАЛЛИФИ АСАРҲОИ ТАЪРИХИЮ-СИЁСӢ БО ВОСИТАИ УНИГРАММАҲОИ РАМЗӢ

Қосимов А.А. –

номзади илмҳои техникӣ, омӯзгори калон, кафедраи барномарезӣ ва низомҳои иттилоотӣ, Донишқадаи политехникии Донишгоҳи техникии Тоҷикистон ба номи академик М.С. Осимӣ, abdunabi.kbtut@mail.ru

Қаюмов А.А. –

магистрант, Академияи миллии илмҳои Тоҷикистон Маркази илмии Хуҷанд, zohirjon96@gmail.com

Муқаррар карда шуд, ки басомади вохӯрии униграммаҳои забони тоҷикӣ дар асарҳои таърихӣ - сиёсӣ ин муайянкунандаи муаллиф аст. Барои муайянкунии муаллифи матн бо басомади вохӯрии ҳарфҳои униграмма ва имконияти истифодабарии таснифгари Усмонов З.Ҷ. таҳқиқот гузаронида шуд. Роҳҳо, мақсад ва имкониятҳои таснифгари Усмонов З.Ҷ. оварда мешаванд. Симои рақамӣ ва фазои ченаки асарҳои таърихӣ-сиёсӣ сохта шудаанд. Бо мақсади ягонагии эҷодиёти муаллиф, қимати ченаки муқоисакунанда муқаррар карда мешаванд, ки дар асоси он ягонагии синфҳои асарҳо муайян мешаванд. γ-таснифгари бузургии тасодуфии фосиладор, ки самаранокии баландро дар муайян кардани муаллифи асарҳои шоирону нависандагони форсу тоҷик нишон дода буданд, барои муайян кардани муаллифи асарҳои таърихӣ-сиёсӣ тафтиш карда мешаванд.

Натиҷаҳое, ки олимони дар ин ҷо ба даст оварданд, фақат барои асарҳои бадеӣ ва иқтисодӣ буданд. Аз ҳамин нуқтаи назар дар ин мақола маълумот оид ба таҳқиқоти муайян кардани муаллифи асарҳои таърихӣ-сиёсӣ бо воситаи басомади вохӯрии униграммаҳои рамзӣ оварда мешавад. Масъалаи шинохти матн, дар асл, аз он рӯзе, ки хат пайдо шуд, ба миён омадааст. Дар муддати тӯлони, он танҳо бо яке аз “рӯя”-ҳои худ, ки дар он зарурати муайян кардани иҷроқунандаи асари

навишташуда ифода меёфт, мавриди таҳлил ва баррасӣ қарор дода мешуд. Минбаъд, пас аз лаҳзаи ихтироъ шудани “китобчопкунӣ”, дар ин масъала, рӯяи нави муҳими дигари он – зарурати муайян кардани муаллифи маводи чопӣ ба миён омад. Дар марҳилаи кунунӣ маҳз ҳамин “рӯя”-и дуюм, мазмуни мундариҷаи асосии ин масъаларо ташкил медиҳад.

Дар кори зерин ба сифати инструменти таҳқиқотшаванда таснифгари Усмонов З.Ҷ. санҷида мешавад, [3-5].

Маълумот оид ба коллексияи матнҳо. Асарҳои муаллифҳои зерин гирифта шуданд ва дар дохили қафс шакли кӯтоҳ карда шудаи онҳо, ки барои ҷойгиркунии ҷадвалҳои 1 ва 2-и поёни лозим буд, оварда шудаанд: А. Дюма (АД) “Граф монте кристо 1” (ГМК1) ва “Граф монте кристо 2” (ГМК2) [6, 7], А. Хуш (АХ) “Шӯриши Усмон 1” (ШУ1) ва “Шӯриши Усмон 2” (ШУ2) [8-10], М. Шакурӣ (МШ) “Садри Бухоро” (СБ) ва “Хуросон аст инчо” (Х) [11, 12], С. Айни (СА) “Ёддоштҳо 1” (Ё1) ва “Ёддоштҳо 2” (Ё2) [13, 14], Ҷ. Иқромӣ (ҶИ) “Дувоздаҳ дарвозаи Бухоро” (ДБ) ва “Он чи аз сар гузашт” (О) [15, 16]. Ба омӯзиши ин масъала ҳамагӣ 5-муаллиф ва 10-асар гирифта шуд.

Таснифгари матнҳо. Ба сифати тавсири рақамии асарҳои таърихӣ-сиёсӣ басомади вохӯрии униграммаҳои рамзӣ дида баромада мешавад. Барои муайянкунии муаллиф як метод – таснифгари матнии Усмонов З.Ҷ. истифода бурда шуд. Моҳияти тавсифи ин метод дар татбиқ ба масъалаҳои илми забоншиносӣ дода мешавад, [3].

Бигзор T_1 ва T_2 – ду матне бошанд, ки қонуни тақсимоти униграммаҳои рамзии онҳо ба намуди ҷадвал дода шуда бошад

$$\begin{aligned} T_i &: 1 \dots k \dots m \\ P^{(i)} &: p_1^{(i)} \dots p_k^{(i)} \dots p_m^{(i)}, \end{aligned} \quad (1)$$

ки дар ин ҷо

$$\sum_{k=1}^m p_k^{(i)} = 1 \quad \text{аст.}$$

Дар ин ифодаҳо k ($k = \overline{1, m}$) - рақами тартибии униграммаи k -юм дар алифбои униграмма, $p_k^{(i)}$ - басомади нисбии вохӯрии униграммаи k -

юм дар матнҳои T_i , $i = 1, 2$ мебошад. Он гоҳ масофаи байни T_1 ва T_2 бо формулаи зерин муайян карда мешавад

$$\rho(T_1, T_2) = \sqrt{\frac{m}{2}} \max_s \left| \sum_{k=1}^s (p_k^{(1)} - p_k^{(2)}) \right|, \quad (2)$$

дар ин ҷо $s = \overline{1, m}$.

Бигзор γ - ягон адади мусбат бошад, матнҳои T_1 ва T_2 γ -якҷинса номида мешаванд, агар

$$\rho(T_1, T_2) \leq \gamma. \quad (3)$$

ва γ -ғайриякҷинса номида мешаванд, агар

$$\rho(T_1, T_2) > \gamma \text{ бошад.} \quad (4)$$

Фарз мекунем, ки коллексияи матнҳо T ба зермаҷмӯъҳои $T^{(j)}$, $j = \overline{1, n}$ тақсим шудааст. Барои қиммати қайдшудаи γ адади \aleph^0 -суммаи ҷуфтҳои якҷинсаи матн, ки ба зермаҷмӯъҳои $T^{(j)}$, $j = \overline{1, n}$, тааллуқ доранд ва адади \aleph^H -суммаи γ -ҷуфтҳои ғайриякҷинса, ки ба зермаҷмӯъҳои гуногун тааллуқ дорад, ҳисоб карда мешавад. Нисбати

$$\eta = \frac{\aleph^0 + \aleph^H}{\aleph}, \quad (5)$$

ки дар ин ҷо \aleph -шумораи умумии ҷуфти матнҳо дар коллексияи T аст, барои қиммати дода шудаи γ самаранокии татбиқи модели математикии (1) – (4) ба таври автоматӣ тақсимкунии коллексияи T ба зерқисми $T^{(j)}$ -ро тавсир мекунад. Дар мақолаҳои [4, 5], барои ҳисоб кардани қимати оптималии $\gamma^{\text{опт}}$, ки барои он самаранокии максималии η барои коллексияи $T = \{T^{(j)}\}$ дастрас мегардад, пешниҳод гардид.

Натиҷаҳо. Алгоритми дар боло зикр шударо истифода бурда, комплекси барномаҳо тартиб дода шуданд ва дар аввал басомади вохӯрии униграммаҳои рамзӣ бе ва бо ҳисобгирии фосила дар

алоҳидагӣ ҳисоб карда, баъдан масофаи байни асарҳо бо формулаи (2) ҳисоб карда шуданд, натиҷаҳо дар ҷадвали 1 ва 2 оварда шудаанд. Аз натиҷаҳои ба даст омада чунин қонуниятро бояд ҷудо кард, ки ду асари як муаллиф якҷинсаанд ва ду асари муаллифашон гуногун ғайриякҷинсаанд.

Ҷадвали 1

Қиммати γ^{opt} барои униграмма бе ҳисобгирии фосола

АД		АХ		МШ		СА		ЧИ	
ГМК1	ГМК2	ШУ1	ШУ2	СБ	Х	Ё1	Ё2	ДБ	О
0,0192									
0,0839	0,0672								
0,0765	0,0705	0,0267							
0,1682	0,1514	0,0843	0,0916						
0,1916	0,1749	0,1077	0,1151	0,0394					
0,0715	0,0704	0,1083	0,1062	0,1910	0,2145				
0,0487	0,0477	0,0844	0,0834	0,1567	0,1802	0,0372			
0,0583	0,0419	0,0642	0,0806	0,1445	0,1679	0,0779	0,0561		
0,0317	0,0404	0,0532	0,0502	0,1374	0,1609	0,0611	0,0383	0,0501	

Бояд қайд кард, ки дар ҳар ду ин ҷадвал дар диагонали асосӣ маълумот оид ба муносибати байни асарҳои як муаллиф, аммо дар дигар ячейкаҳо маълумот оид ба муносибати байни асарҳои муаллифашон гуногун оварда шудаанд.

Барои муайянкунии муаллиф диапазони қиммати мувофиқи γ бо воситаи униграмма бе ҳисобгирии фосола баробари $\gamma = [0,0317; 0,0371]$ шуд. Дар ин ҳолат бо нобаробарии зерин

$$\rho(T_1, T_2) \leq [0,0317; 0,0371] \quad (6)$$

- якҷинсагии ҷуфти асарҳо, аммо бо муқобили нобаробари,

$$\rho(T_1, T_2) > [0,0317; 0,0371), \quad (7)$$

- бо ғайриҷинсагии асарҳо мувофиқат мекунад. Ин қоидаро ба қатори ададҳои ҷадвали 1 татбиқ намоем, нобаробарии (6) дар 10 ячейка риоя намешавад, аммо нобаробарии (7) фақат дар 2 ячейка риоя намешавад. Ба ҳолати зерин, таъсири метод бо формулаи (5) ҳисоб карда шуд, ки баробари $\eta = 93\%$ аст.

Ҷадвали 2

Қиммати γ^{opt} барои униграмма бо ҳисобгирии фосила

АД		АХ		МШ		СА		ҶИ	
ГМК1	ГМК2	ШУ1	ШУ2	СБ	Х	Ё1	Ё2	ДБ	О
0,0229									
0,0743	0,0670								
0,0772	0,0727	0,0184							
0,1074	0,1016	0,0568	0,0585						
0,1386	0,1328	0,0881	0,0897	0,0534					
0,0644	0,0562	0,0764	0,0798	0,1321	0,1633				
0,0496	0,0415	0,0617	0,0651	0,1085	0,1400	0,0263			
0,0302	0,0341	0,0785	0,0813	0,1029	0,1275	0,0584	0,0503		
0,0551	0,0474	0,0402	0,0413	0,0971	0,1283	0,0419	0,0271	0,0581	

Таҳлили ҷадвали 2 нишон медиҳад, ки бо воситаи униграмма бо ҳисобгирии фосила, таснифгар ҳиссиёти баландтарро дора аст. Барои

муайянкунии муаллиф диапазони қиммати мувофиқи γ бо воситаи униграмма бо ҳисобгирии фосола баробари $\gamma = [0,0271; 0,0581)$ шуд. Дар ин ҳолат бо нобаробарии зерин

$$\rho(T_1, T_2) \leq [0,0271; 0,0581) \quad (8)$$

- якҷинсагии ҷуфти асарҳо, аммо бо муқобили нобаробари,

$$\rho(T_1, T_2) > [0,0271; 0,0581), \quad (9)$$

- бо ғайриҷинсагии асарҳо мувофиқат мекунад. Ин қоидаро ба қатори ададҳои қадвали 2 татбиқ намоем, нобаробарии (8) дар 10 ячейка рӯя намешавад, аммо нобаробарии (9) бошад, фақат дар 2 ячейка рӯя намешавад. Ба ҳолати зерин, таъсирнокии метод бо формулаи (5) ҳисоб карда шуд, ки баробари $\eta = 96\%$ аст.

Хулоса. Аз маълумотҳои ҳангоми таҳқиқот ба даст омада, ба чунин хулосаҳо омадан мумкин аст, ки

- униграммаи рамзӣ дар масъалаи муайянкунии муаллифи матни таърихӣ-сиёсӣ ба сифати тавсифҳои миқдорӣ комилан қобили қабул мебошанд;

- ба ҳисобгирии фосола дар униграммаҳо саҳеҳии таснифотро баланд мебардорад;

- таснифгари Усмонов З.Ҷ. (1) – (5) дараҷаи кифоя калони муайянкунии муаллифи матнҳои таърихӣ-сиёсиро нишон медиҳад.

ЛИТЕРАТУРА

1. Усманов З.Д., Косимов А.А. Частотность униграмм в таджикской литературе.– Доклады Академии наук Республики Таджикистан, 2016, т.59, № 1-2, с. 28-32.

2. Романов А.С., Шелупанов А.А., Мещеряков Р.В. Разработка и исследование математических моделей, методик и программных средств информационных процессов при идентификации автора текста.– Томск: -В-Спектр, 2011, 188 с.

3. Усманов З.Д. N-граммы в распознавании однородных текстов.– Материалы 20 научно-практического семинара "Новые

информационные технологии в автоматизированных системах".– Москва, 2017, с. 52-54.

4. Усманов З.Д. Классификатор дискретных случайных величин.– Доклады Академии наук Республики Таджикистан. 2017, т.60, № 7-8, с. 291-300.

5. Усманов З.Д. Алгоритм настройки кластеризатора дискретных случайных величин.– Доклады Академии наук Республики Таджикистан, 2017, т.60, № 9, с. 392-397.

6. Александр Дюма Граф монте кристо. 1844, 520 с.

7. Алӣ Хуш Шӯриши Усмон . Душанбе, 2008, 128 с.

8. Муҳаммадҷон Шакурӣ Садри Бухоро. Душанбе, 2005, 330 с.

9. Муҳаммадҷон Шакурӣ Хуросон аст инчо.

10. Садриддин. Айнӣ Ёддоштҳо. 1990, 352 с.

11. Ҷалол. Икромӣ Дувоздаҳ дарвозаи Бухоро. 1969, 474 с.

12. Ҷалол. Икромӣ Он чи аз сар гузашт. Душанбе, 2009, 360 с.