

Let data sing the Uyghur Twelve Muqam: A text mining on lyrics

Kai, Limai

kalima@ruc.edu.cn
School of Information Resource Management, Research Center for Digital Humanities, Renmin University of China, China, People's Republic of

Yang, Zekun

zekunyang@ruc.edu.cn
School of Information Resource Management, Research Center for Digital Humanities, Renmin University of China, China, People's Republic of

Feng, Huiling

fhl@ruc.edu.cn
School of Information Resource Management, Research Center for Digital Humanities, Renmin University of China, China, People's Republic of

Liu, Yuenan

liuyuenan@ruc.edu.cn
School of Information Resource Management, Research Center for Digital Humanities, Renmin University of China, China, People's Republic of

Liang, Jihong

liangjihong@ruc.edu.cn
School of Information Resource Management, Research Center for Digital Humanities, Renmin University of China, China, People's Republic of

Yang, Huilin

yanghuilin@ruc.edu.cn
Centre for European Studies/Centre for EU Studies, Renmin University of China, China, People's Republic of

Yan, Jin

yanjin@ruc.edu.cn
Centre for European Studies/Centre for EU Studies, Renmin University of China, China, People's Republic of

Huang, Yanfen

huangyf@ruc.edu.cn
Centre for European Studies/Centre for EU Studies, Renmin University of China, China, People's Republic of

Guan, Kongwen

gkw9122@ruc.edu.cn
Centre for European Studies/Centre for EU Studies, Renmin University of China, China, People's Republic of

Introduction

Originated in the 4th century, Uyghur Twelve Muqam is an artistic treasure created by the Uyghur people, which integrates songs, dances, and instrumental performances (Han & Xiao, 2017a; Harris, 2008; Tursun et al., 2019). The Uyghur Muqam of Xinjiang was designated as an Intangible Heritage of Humanity by UNESCO in 2005. Previous research on Uyghur Twelve Muqam mainly focuses on its history, origin, rhythm, tone, structure, notation, dance, lyrics, and characters (Anderson, 2012; Han & Xiao, 2017a; Han & Xiao, 2017b; Imin et al., 2019; Light, 1998). However, digital humanities methodology has not been applied to study the Uyghur Twelve Muqam.

Thus, we start the project "Let data sing the Uyghur Twelve Muqam", which applies text mining methods to gain deeper insights into Muqam lyrics. This paper demonstrates the preliminary results, which discover new insights beyond existing qualitative research on Muqam, including new Muqam poem topics, sentiment styles of individual poets, and topic similarity among poets.

Method

We conduct topic modeling, sentiment analysis, and social network analysis to analyze Muqam lyrics. The techniques used and the common techniques shared among the three text mining approaches are shown in Figure 1.

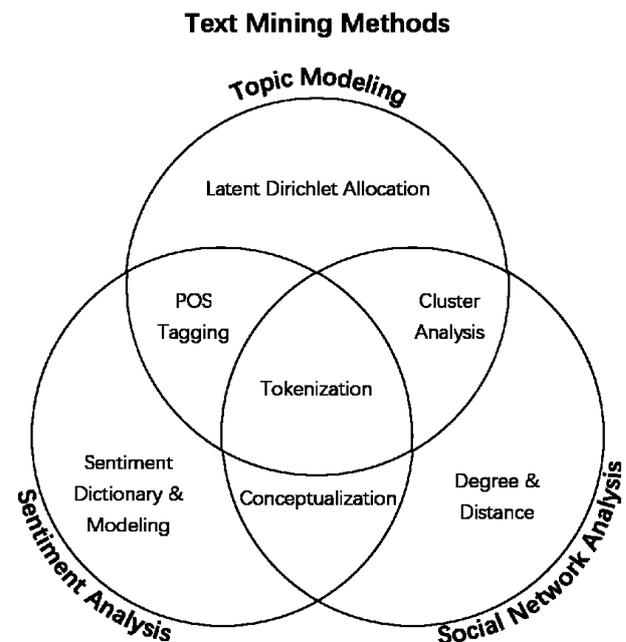


Figure 1: Text Mining Methods Used in the Project

Corpus

The Chinese-translated Muqam lyrics corpus is gathered from two Uyghur Twelve Muqam books (Dawamat, 1997; Li & Meng, 2008). The corpus contains twelve sets of Muqam, each composed of three parts, each part is further separated into two to thirteen smaller parts, and each smaller part contains one to five poems. Figure 2 shows the structure of a Muqam.



>Figure 2: The Structure of the Muqam

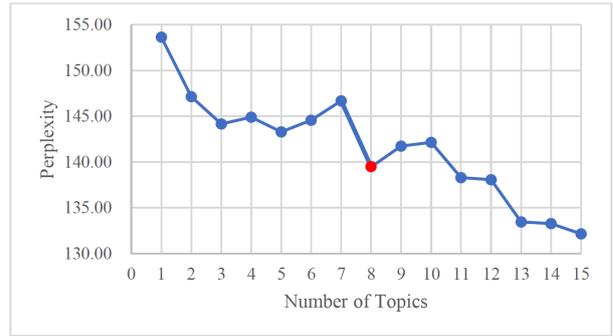


Figure 3: Perplexity of LDA

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
Disaster	Appearance	Love	Wine	Kingdom	Cry	Garden	Faith
disaster	cherry lips	lover	lover	Allah	heart	fairy	Allah
world	face	pain	wine	flower	hair	home	nightingale
sun	girl	love	beggar	lover	pain	goddess	lover
love	hair	garden	pub	monarch	lover	paradise	rose
pain	eyes	nightingale	soul	slave	love	juniper	garden
destiny	visage	slender	thirst	nightingale	blood tears	miss	faithful
society	nightingale	promise	Allah	beauty	body	valerian	face
suffer	veil	flowerbed	winecup	beggar	tears	cherry lips	precious
enemy	everybody	destiny	secret	king	pearl	musk	king
friend	body	tryst	destiny	binocular	eyes	flowerbed	sadness
man's world	wish	misery	moth	heaven	beauty	flower	lamentation
joy	haggard	wish	world	love	soul	hair	man's world
people	sun	tender	tavern	alley	eyelash	oasis	care
infatuation	vellus hair	wonderful	wine glasses	wealth	flower	black eyes	piety
scar	soul	robber	believer	world	lips	garden	sorrow

Figure 4: Topic Modeling of Uyghur Twelve Muqam

Preliminary Experimental Results

Next, we demonstrate the preliminary results of mining Uyghur Twelve Muqam lyrics using topic modeling, sentiment analysis, and social network analysis.

Topic Modeling

We apply the statistical topic model Latent Dirichlet Allocation (LDA) to extract topics of semantically related words in Muqam poems, instead of manually identifying topics. We first preprocess the Muqam text using tokenization and POS tagging methods in the Python package Jieba, and keep only the nouns and adjectives in the poems for topic modeling. Then, we use the Python package scikit-learn to conduct LDA. By calculating the elbow point based on perplexity (shown in Figure 3), we set the number of topics of LDA to eight. Figure 4 shows the eight topics identified. Among the eight topics, four are already identified by previous research, which are love, cry, disaster, and faith (Zhou, 1995). The other four topics, including appearance, wine, kingdom, and garden, are newly discovered by our research.

Sentiment Analysis

Sentiment analysis identifies emotion from texts. Previous Muqam research studies the sentiment difference among different parts of Muqam, while the sentiment difference among different poets has not been investigated. Thus, we employ the Python package baidu-ai to directly calculate the sentiment of each poem and analyze the sentiment distribution of poets. The number and percentage of positive poems written by each of the Muqam poets are shown in Figure 5. We could observe that among the 41 poets, 31 prefer to write comedies, and only five write more tragedies.

Poet	Number of Positive Poems	Percentage of Positive Poems	Poet	Number of Positive Poems	Percentage of Positive Poems
Näwä'i	33	80.49%	Häfiz	2	66.67%
Mäšrab	10	90.91%	Xaräzmi	2	100.00%
Näwbäti	9	100.00%	Šäyfi	2	100.00%
Mäšhuri	7	77.78%	Särayi	2	100.00%
Näsimi	6	85.71%	Niyazi*	2	100.00%
Ärsi	5	83.33%	Firaqi	2	66.67%
Huwäyda	4	80.00%	Šadiqi*	2	100.00%
Zäliili	4	100.00%	Atayi	2	66.67%
Futuhi	4	100.00%	Zumurräd	1	100.00%
Bilal Nazim	4	80.00%	Fuzuli	1	100.00%
Mähzun	3	33.33%	Nizali*	1	100.00%
Gäda'i	3	100.00%	Räšidi	1	50.00%
Ayazi	3	75.00%	Lutfi	1	33.33%
Babir	3	75.00%	Riža'i	1	100.00%
Zühuri	3	60.00%	Näsimi	1	100.00%
Gumnum	3	100.00%	Älib'i*	1	100.00%
HäfizXaräz	3	75.00%	Nakis*	1	100.00%
Qäländär	3	50.00%	Mähzal*	1	100.00%
Wäfa'i	2	100.00%	Šalihi	0	0.00%
Šäyid*	2	100.00%	Näfis	0	0.00%
Säkkari	2	50.00%	Säburi	0	0.00%

*The asterisk mark indicates that the name is translated by the author according to Transliteration used from post-1978 Uyghur Arabic Script

Figure 5: Sentiment Distribution of Muqam Poets

Social Network Analysis

Social network analysis measures relationships between social actors. While previous research focuses on analyzing Muqam poets individually (Zhou, 2006), social network analysis could discover the topic similarity relationship between multiple poets. We select nouns that appear more than ten times in Muqam poems, calculate the noun co-occurrence between poets, and visualize the co-occurrence relationship using a social network via Gephi. Figure 6 shows the noun co-occurrence network between Muqam poets. The thicker the lines in the graph, the greater number of nouns shared by the two poets. We could observe that many poets share similar nouns and topics in their poems.

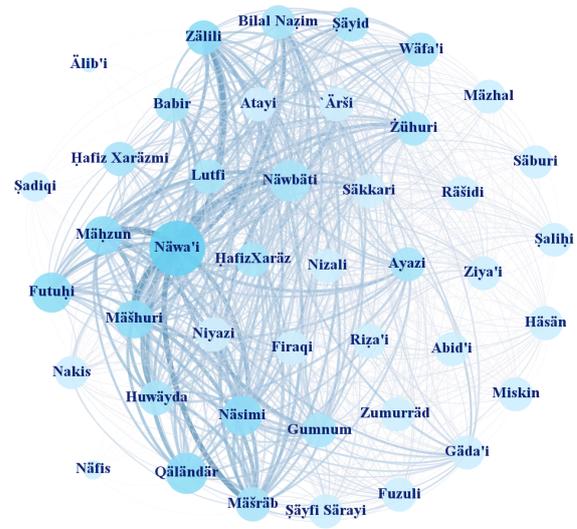


Figure 6: Noun Co-Occurrence Graph Between Muqam Poets

Conclusion and Future Work

In this paper, we introduce the preliminary results of our project “Let data sing the Uyghur Twelve Muqam”. Based on text mining, we discover Muqam poem topics, sentiment styles of individual poets, and topic similarity among poets that have not been discovered by previous qualitative Muqam research. In the next phase of the project, we will develop natural language processing tools for Uyghur, and use these self-developed tools to analyze the Uyghur version of Muqam poems.

Zekun Yang is the corresponding author and can be contacted at zekunyang@ruc.edu.cn

Bibliography

- Anderson, Elise (2012): “The Construction of Ämännisä Khan as a Uyghur Musical Culture Hero”, in: *Asian Music* 43,1: 64-90.
- Dawamat, Tumor (Eds.) / Xinjiang Uyghur Autonomous Region Twelve Muqam Research Society / Xinjiang Uyghur Autonomous Region Classical Literature Research Society (1997): *Uyghur Twelve Muqam*. China Encyclopedia Publishing House.
- Han, Yunxia / Xiao, Xuejun (2017a): “Summary of The Twelve Muqam study (1949-2015) (1) (Chinese)” in: *Journal of Xinjiang Arts University* 15, 1:8-25.
- Han, Yunxia / Xiao, Xuejun (2017b): “Summary of The Twelve Muqam study (1949-2015) (2) (Chinese)” in: *Journal of Xinjiang Arts University* 15, 2:1-10.
- Harris, Rachel (2008): *The making of a musical canon in Chinese Central Asia: the Uyghur Twelve Muqam*. Ashgate Publishing, Ltd.
- Imin, Nurbiye / Tudaji, Rehemitulla / Juma, Osman (2019): “Study on China’s Uyghur Twelve Muqam in the Past 40 Years of Reform and Opening up”, in: *Third International Conference on Art Studies: Science, Experience, Education*, Moscow, Russia, October 2019 :78-80. DOI: 10.2991/icassee-19.2019.15

Li, Dongming / Meng, Xing (Eds.) (2008): *Twelve Muqam Lyrics (1st ed.)*. Xinjiang People's Publishing House.

Light, Nathan (1998): *Slippery paths: the performance and canonization of Turkic literature and Uyghur Muqam song in Islam and modernity*. Ph.D. Thesis, Indiana University.

Tursun, Guzelnur / Tudaji, Rehemitulla / Juma, Osman (2019): "Overview of China's Uyghur Twelve Muqam in the Past 40 Years of Reform and Opening up". in: *Third International Conference on Art Studies: Science, Experience, Education*, Moscow, Russia, October 2019 :78-80. DOI: 10.2991/icassee-19.2019.35

Zhou, Ji (1995): "Ten topics in the Uyghur Twelve Muqam (Continued 1) (Chinese)", in: *Journal of Xinjiang Normal University (Edition of Philosophy and Social Science)*, 2: 14–25.

Zhou, Ji (2006): "A review of Uyghur Muqam research in the 20th century (Chinese)". in: *Journal of Xinjiang Arts University*, 4, 11–17.