



RECOMMENDATION APPROACH OF ENGLISH SONGS TITLE BASED ON LATENT DIRICHLET ALLOCATION APPLIED ON LYRICS

(แนวทางการแนะนำชื่อเพลงภาษาอังกฤษ
โดยอิงจากการจัดสรร LATENT DIRICHLET กับเนื้อเพลง)

Date of Conference

20 Feb, 2019

DOI: 10.1109/ICECCT.2019.8869198

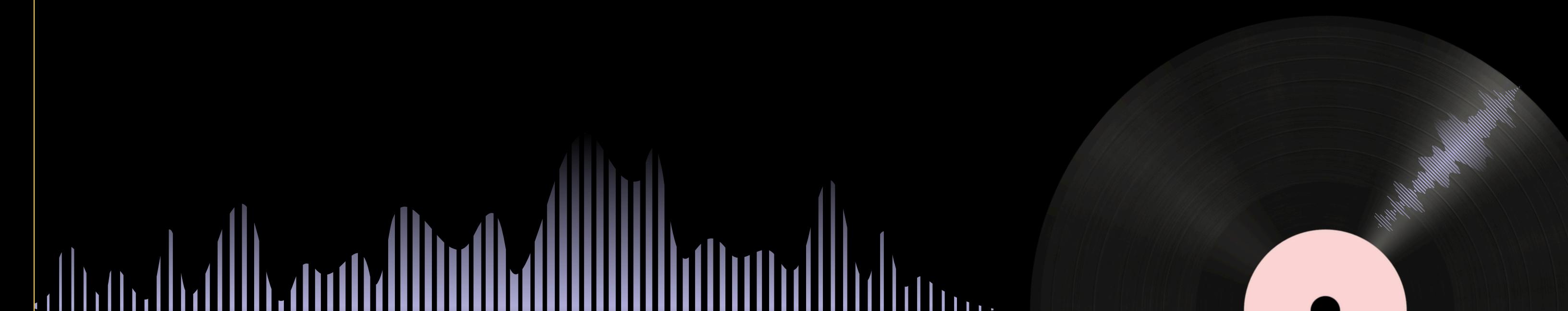
Published in: 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)



ABSTRACT



เพลงถือเป็นส่วนสำคัญที่สามารถส่งผลโดยตรงต่อความรู้สึกของผู้ฟัง การเลือกซื้อเพลงที่เหมาะสมเป็นปัจจัยที่สำคัญมาก สำหรับผู้ฟัง เนื่องจากซื้อเพลงสร้างความประทับใจแรก การกำหนดชื่อเพลงที่เหมาะสมจึงเป็นงานที่ท้าทาย เนื้อเพลงเป็นส่วนที่มีอิทธิพลมากที่สุดนอกเหนือจากองค์ประกอบอื่น ๆ เช่น ทำนอง จังหวะ นักร้อง หรือแนวเพลง งานวิจัยนี้ใช้เทคโนโลยี LATENT DIRICHLET ALLOCATION (LDA) เพื่อค้นหา TOPIC ที่ซ่อนอยู่ในเนื้อเพลง โดยผลลัพธ์จาก LDA จะให้แบบจำลองความสำคัญของคำ ซึ่งจะถูกนำไปใช้ในการบันการประมาณค่าเพื่อสร้างชื่อเพลงที่เหมาะสม การทดลองได้ดำเนินการกับชุดข้อมูลเนื้อเพลงภาษาอังกฤษจำนวนกว่า 200 เพลงที่มีความหลากหลายในแนวเพลง และวิธีการนี้สามารถประเมินผลได้โดยการเปรียบเทียบกับชื่อเพลงเดิม ซึ่งใช้หลักการเดียวกับระบบแนะนำ (RECOMMENDATION SYSTEM)



INTRODUCTION

ในงานวิจัยนี้มีการนำทำ LDA ซึ่งเป็นเทคนิคการทำ Topic Modeling มาใช้เพื่อค้นหาคำสำคัญที่ซ่อนอยู่ในเนื้อเพลง โดย LDA จะวิเคราะห์เอกสารหรือข้อความผ่านกระบวนการทางสถิติเพื่อหา “Topic” ที่ซ่อนอยู่ และแบ่งนำชื่อเพลงที่เหมาะสมจากคำสำคัญที่ได้ โดยมีโครงสร้างของบทความดังนี้



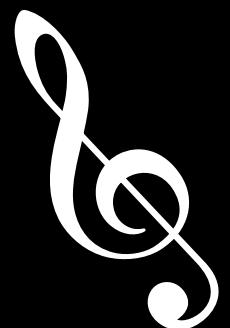
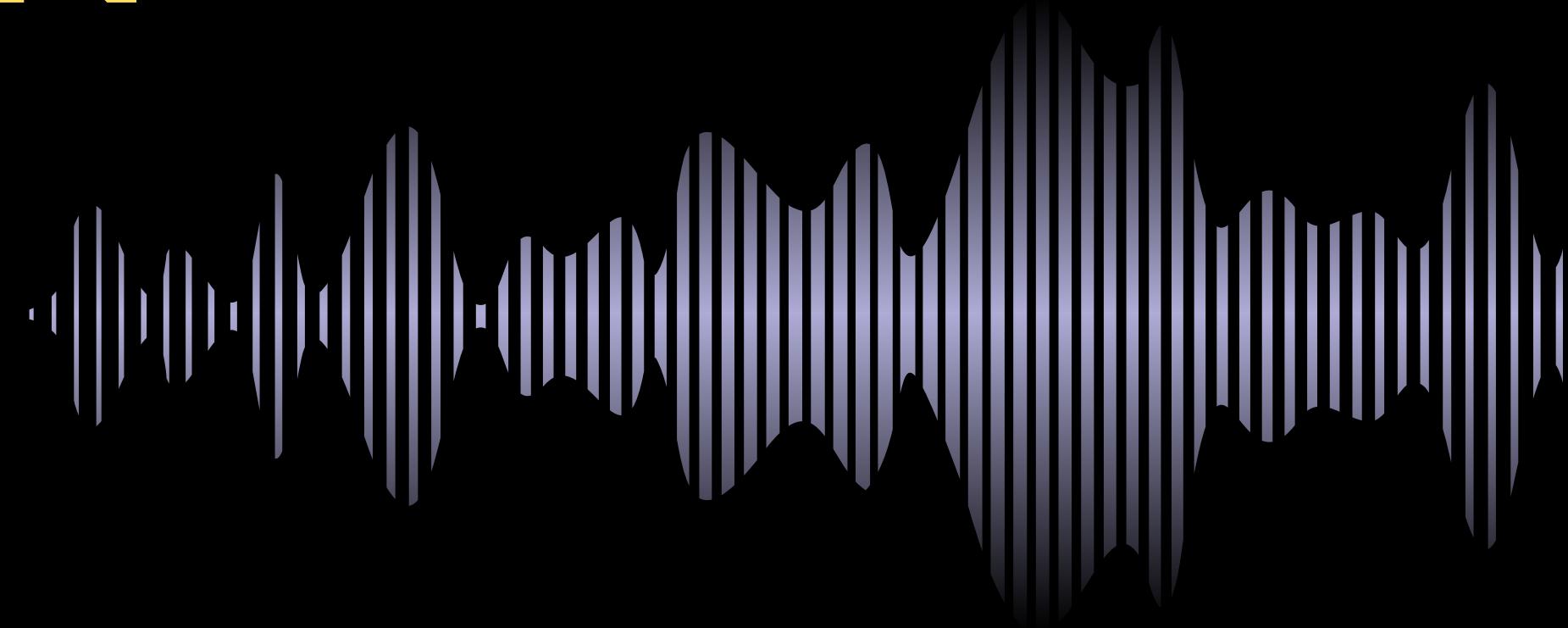
I : Proposed Approach



II : Experiments



III : Conclusion





I.

I. PROPOSED APPROACH



[Home](#)

[Music](#)

[About](#)

[Contact](#)

PROPOSED APPROACH

แบ่งออกเป็น 5 ส่วน

- Data Collection

รวบรวมเพลงภาษาอังกฤษและเนื้อเพลงต้นฉบับที่
รวบรวมจากแหล่งข้อมูลบนเว็บและข้อมูลอัลบัมต่างๆ
ซึ่งประกอบด้วยข้อมูลเนื้อเพลง นักร้อง และนักแต่ง
เพลง 200 เพลง

- Data Processing

- ลบคำที่ไม่จำเป็น เช่น " [verse] ", " [chorus] "
และตัวเลขที่แสดงการซ้ำ
- แก้ไขคำย่อในทุกเพลงในชุดข้อมูล เช่น " isn't " เป็น " is not "
" didn't " เป็น " did not "

[Verse 1]

This ain't a song for the broken-hearted
No silent prayer for the faith-departed
I ain't gonna be just a face in the crowd
You're gonna hear my voice
When I shout it out loud

[Chorus]

It's my life
It's now or never
I ain't gonna live forever
I just want to live while I'm alive
My heart is like an open highway
Like Frankie said, "I did it my way"
I just wanna live while I'm alive
It's my life.

This is not a song for the broken-hearted
No silent prayer for the faith-departed
I am not gonna be just a face in the
crowd
You are gonna hear my voice
When I shout it out loud

It is my life

It is now or never

I am not gonna live forever

I just want to live while I am alive

My heart is like an open highway
Like Frankie said, I did it my way
I just wanna live while I am alive
It is my life.

(a)

(b)

รูปภาพจาก : <https://ieeexplore.ieee.org/document/8869198>

PROPOSED APPROACH

- Applying LDA for Finding Topic

ใช้ Latent Dirichlet Allocation (LDA) ซึ่งเป็นอัลกอริธึมการเรียนรู้แบบไม่มีผู้สอน ค้นหาหัวข้อของเนื้อเพลง หาคำและค่าความน่าจะเป็นที่เกี่ยวข้องซึ่งจะส่งต่อไปยังกระบวนการประมาณค่าถัดไป

- Estimation Process

- ใช้ค่าความน่าจะเป็นที่ได้จาก LDA เพื่อหาคำสำคัญ 3 คำที่มีค่าความน่าจะเป็นสูงสุด

- Song Title Recommendation

ใช้กฎไวยากรณ์สร้างประโยคที่มีความหมายจากคำสำคัญที่ได้ สร้างชื่อเพลงที่แนะนำโดยใช้คำสำคัญที่ได้มา

Estimation Process

Input: Topic words ($wr_1, wr_2, wr_3, \dots, wr_n$) and their associated probabilistic values ($v_1, v_2, v_3, \dots, v_n$), where n is number of words

Output: Essential words ($ew_1, ew_2, ew_3 \dots$) list to determine the song title.

Steps:

1. Scan the topic words ($wr_1, wr_2, wr_3 \dots wr_n$) and their associated probabilistic values ($v_1, v_2, v_3 \dots v_n$), where n is number of words.
2. Find $\{maxVal\}$: $\{\max(v_1, v_2, v_3, \dots, v_n)\}$
3. Find $\{secMaxVal\}$: $\{\max(\{v_1, v_2, v_3, \dots, v_n\} - \{maxVal\})\}$
4. Find $\{thirdMaxVal\}$: $\{\max(\{v_1, v_2, v_3 \dots v_n\} - \{maxVal\} - \{secMaxVal\})\}$.
5. Find topic words having ($maxVal, SecMaxVal, thirdMaxVal$)
6. List the topics words as essential words (ew_1, ew_2, ew_3)



II.

II. EXPERIMENTS



[Home](#)

[Music](#)

[About](#)

[Contact](#)

EXPERIMENTS

ดำเนินการตามลำดับขั้นตอนของวิธีการที่นำเสนอไว้

- Data Collection

รวบรวมชุดข้อมูลเพลงภาษาอังกฤษจำนวน 200 เพลงซึ่งมีเนื้อเพลงของแต่ละเพลง เก็บชื่อเพลงเดิม ไว้[1] และคุณลักษณะและคุณสมบัติบางประการของ ชุดข้อมูล[2]

- Data Processing

การแยกคำที่ไม่จำเป็นออกจากเนื้อเพลงที่ ใช้เครื่องมือ Natural Language Toolkit (NLTK) และไลบรารี gensim ใน Python เพื่อกรองคำชำ្លៃចានจากชุดข้อมูล ผลลัพธ์จากขั้นตอนนี้[2]คือ เนื้อเพลงที่ปรับปรุงแล้ว โดยมีการขยายคำย่อและลบคำชำ្លៃ

[1]

Song ID	Original Song Title
Song1	A PILLOW OF TEARS
Song2	SEE YOU AGAIN
Song3	WHAT DID I DO WRONG
Song4	HAPPY
Song5	IT'S MY LIFE
Song6	I HAVE BEEN HERE BEFORE
Song7	I JUST WANNA GOOD TIME
Song8	I NEED YOU BESIDE ME
Song9	I'M NOT LOOKING FOR YOUR LOVE
Song10	HOTEL CALIFORNIA

[2]

Properties	Values
No. of Songs	200
Avg. Words (per song)	128.385
Total Words	25677
Avg. Noun	33.74
Avg. Pronoun	25.715
Avg. Adjective	6.52
Avg. Verb	33.555
Avg. Adverb	11.190
Avg. Preposition	14.740
Avg. Conjunction	4.020
Avg. interjection	0.165

EXPERIMENTS

- Applying LDA for Finding Topic

ใช้อัลกอริธึม LDA แบบปรับแต่ง ซึ่งไม่ใช้ stemming เพื่อรักษาคำเดิมของเนื้อเพลง

ตัวอย่างการทดลองกับเพลง "Happy" โดย Pharrell Williams ได้คำสำคัญดังนี้

```
Happy --0.102; along --0.088; clap --0.088;  
like --0.071; feel --0.067
```

รูปภาพจาก : <https://ieeexplore.ieee.org/document/8869198>

ค่าความน่าจะเป็นเหล่านี้จะถูกส่งไปยังกระบวนการ
ประมาณค่าในขั้นตอนถัดไป

- Estimation Process

กำหนดจำนวนคำขึ้นต่อที่จำเป็นในการสร้างประโยคที่เรียบง่าย โดยใช้ไวยากรณ์ ต้องใช้คำอย่างน้อยสามคำเพื่อสร้างประโยคที่มีความหมาย ใช้ค่าสูงสุด 3 อันดับแรกจาก LDA เพื่อหาคำสำคัญ

ตัวอย่างจากเพลง "Happy":

- **maxVal** = 0.102 → **happy**
- **secMaxVal** = 0.088 → **clap**
- **thirdMaxVal** = 0.088 → **along**

- Song Title Recommendation

แนะนำชื่อเพลงโดยใช้คำสำคัญที่ได้มาจากการ
ประเมินที่ได้มา ในตัวอย่างที่กล่าวถึงข้างต้น ผลลัพธ์ที่
ได้คือ "Be Happy and Clap Along" หรือ "Clap
Along and Be Happy"



III.

III. CONCLUSION



[Home](#)

[Music](#)

[About](#)

[Contact](#)

CONCLUSION

งานวิจัยนี้ศึกษาความเป็นไปได้ในการใช้ LDA เพื่อสร้างชื่อเพลงโดยวิเคราะห์ความหมายของเนื้อเพลง ผลการทดลองแสดงให้เห็นว่าการใช้ LDA ให้ผลลัพธ์ที่น่าพึงพอใจเมื่อนำไปประยุกต์กับชุดข้อมูลเนื้อเพลงภาษาอังกฤษที่จัดทำขึ้นสำหรับการทดลองนี้

วิธีการนี้นับเป็นงานด้านการประมวลผลภาษาธรรมชาติ (NLP) ที่มุ่งเน้นเฉพาะเนื้อเพลง เนื่องจากใช้ข้อมูลจากเนื้อเพลงเท่านั้น

สำหรับงานวิจัยในอนาคต อาจใช้แบบจำลองหัวข้อที่ซับซ้อนกว่า หรือใช้ LDA ในรูปแบบที่ปรับปรุงแล้ว เพื่อให้ได้ผลลัพธ์ที่แม่นยำยิ่งขึ้น



REFERENCE

R. Hossain, M. R. Kabir Rasel Sarker, M. Mimo, A. Al Marouf and B. Pandey, "Recommendation Approach of English Songs Title Based on Latent Dirichlet Allocation applied on Lyrics," 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2019, pp. 1-4, doi: 10.1109/ICECCT.2019.8869198.
<https://ieeexplore.ieee.org/document/8869198>



THANK
YOU

