

Automatic Survey Understanding

(A NLP & AI Case Study)



George Brova

CTO at Migacore Technologies

george@brova.com



Ling Zhang

Software Engineer at Aiden.ai

lz@ling.nz

About George



George Brova

CTO

MIGACORE

- **Migacore Technologies** uses NLP and AI to forecast travel patterns.
- **Palantir, Google, Groupon**
- Research: Graph Mining + applied ML at **Univeristy of Illinois**

About Ling



Ling Zhang

Software Engineer



- **aiden.ai is an AI powered virtual colleague for marketers.**
- Ex-founder @ Entrepreneur First 8.
- Palantir, Facebook
- NLP, Speech Recognition, Machine Learning

Talk Overview

Part 1:

Problem and Solution

Part 2:

Preprocess the Data

Part 3:

Machine Learning & Analysis

Part 4:

Results

Part 1: Problem and Solution

The Problem

That go on...

Required questions are denoted by an *

1. *Please rate the timeliness of order delivery.

Likert	1	2	3	4	5	6	7	8	9	10	Response ID	Start/End
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>

2. *Please rate the completeness of delivery.

Likert	1	2	3	4	5	6	7	8	9	10	Response ID	Start/End
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>

3. *Please rate the availability of product delivery items.

Likert	1	2	3	4	5	6	7	8	9	10	Response ID	Start/End
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>

4. *Please rate the ease for tracking your order.

Likert	1	2	3	4	5	6	7	8	9	10	Response ID	Start/End
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>

5. *Please rate the condition of the product(s) received.

Likert	1	2	3	4	5	6	7	8	9	10	Response ID	Start/End
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>

6. *Please rate how well the packaging protected your items.

Likert	1	2	3	4	5	6	7	8	9	10	Response ID	Start/End
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>

7. *Please rate the accuracy of your order (do you receive the correct products, sizes, etc.).

Likert	1	2	3	4	5	6	7	8	9	10	Response ID	Start/End
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>

8. *Please rate the completeness of your order (do you receive all the items in your order).

Likert	1	2	3	4	5	6	7	8	9	10	Response ID	Start/End
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>

9. *Please rate the product prices on NextDay.com.

Likert	1	2	3	4	5	6	7	8	9	10	Response ID	Start/End
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>

10. *Please rate the competitiveness of product prices on NextDay.com.

Likert	1	2	3	4	5	6	7	8	9	10	Response ID	Start/End
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>

11. *Please rate the shipping costs on NextDay.com.

Likert	1	2	3	4	5	6	7	8	9	10	Response ID	Start/End
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>

12. *Please rate the reasonableness of shipping costs on NextDay.com.

Likert	1	2	3	4	5	6	7	8	9	10	Response ID	Start/End
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>

13. *Please rate the variety of shipping options on NextDay.com.

Likert	1	2	3	4	5	6	7	8	9	10	Response ID	Start/End
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="radio"/>

And on...

Required questions are identical to an ¹

1. "Please rate the ease of order delivery."	1-5	5	4	3	2	1	Start timer
2. "Please rate the speed of delivery."	1-5	5	4	3	2	1	Start timer
3. "Please rate the availability of desired delivery times."	1-5	5	4	3	2	1	Start timer
4. "Please rate the tools for tracking your order."	1-5	5	4	3	2	1	Start timer
5. "Please rate the condition of the purchased item received."	1-5	5	4	3	2	1	Start timer
6. "Please rate how well the packaging protected your items."	1-5	5	4	3	2	1	Start timer
7. "Please rate the clarity of your order (did you receive the correct products, sizes, etc.)."	1-5	5	4	3	2	1	Start timer
8. "Please rate the completeness of your order (did you receive all the items in your order)."	1-5	5	4	3	2	1	Start timer
9. "Please rate the product prices on BestBuy.com."	1-5	5	4	3	2	1	Start timer
10. "Please rate the competitiveness of product prices on BestBuy.com."	1-5	5	4	3	2	1	Start timer
11. "Please rate the shipping costs on BestBuy.com."	1-5	5	4	3	2	1	Start timer
12. "Please rate the responsiveness of shipping costs on BestBuy.com."	1-5	5	4	3	2	1	Start timer
13. "Please rate the variety of shipping options on BestBuy.com."	1-5	5	4	3	2	1	Start timer

14. "Please rate the variety of shipping options on BestBuy.com."	1-5	5	4	3	2	1	Start timer
14. "Please rate the frequency of email communication regarding the order."	1-5	5	4	3	2	1	Start timer
15. "Please rate the content of email communication regarding the order."	1-5	5	4	3	2	1	Start timer
16. "Please rate the helpfulness of email communication regarding the order."	1-5	5	4	3	2	1	Start timer
17. "Please rate the timeliness of email communication regarding the order."	1-5	5	4	3	2	1	Start timer
18. "What is your overall satisfaction with the entire purchase experience?"	1-5 Very Dissatisfied	5	4	3	2	1 Very Satisfied	Start timer
19. "How well does the purchase experience meet your expectations ?"	1-5 Very Short	5	4	3	2	1 Very Long	Start timer
20. "How does the experience compare to your idea of an ideal purchase experience ?"	1-5 Very Close	5	4	3	2	1 Very Far	Start timer
21. "How likely are you to return to BestBuy.com?"	1-5 Very Unlikely	5	4	3	2	1 Very Likely	Start timer
22. "How likely are you to recommend BestBuy.com to someone else?"	1-5 Very Unlikely	5	4	3	2	1 Very Likely	Start timer
23. "How likely are you to prefer Best Buy over other retailers after this purchase experience?"	1-5 Very Unlikely	5	4	3	2	1 Very Likely	Start timer
24. "How likely are you to make an online purchase from BestBuy.com in the future?"	1-5 Very Unlikely	5	4	3	2	1 Very Likely	Start timer
25. "How likely are you to purchase from one of Best Buy's retail stores in the future?"	1-5 Very Unlikely	5	4	3	2	1 Very Likely	Start timer

And on some more...

Required questions are denoted by an *

1. * Please rate the timeliness of order delivery.

2. * Please rate the convenience of delivery.

3. * Please rate the availability of in-stock delivery items.

4. * Please rate the needs for tracking your order.

5. * Please rate the condition of the product(s) when received.

6. * Please rate how well the packaging protected your items.

7. * Please rate the accuracy of your order (do you receive all the items, products, etc.).

8. * Please rate the completeness of your order (do you receive all the items in your order).

9. * Please rate the product quality on BestBuy.com.

10. * Please rate the competitiveness of product prices on BestBuy.com.

11. * Please rate the shipping costs on BestBuy.com.

12. * Please rate the reasonableness of shipping costs on BestBuy.com.

13. * Please rate the quality of shipping options on BestBuy.com.

14. * Please rate the quality of shipping options on BestBuy.com.

15. * Please rate the quality of shipping options on BestBuy.com.

16. * Please rate the frequency of email communication regarding this order.

17. * Please rate the timeliness of email communication regarding this order.

18. * Please rate the helpfulness of email communication regarding this order.

19. * Please rate the timeliness of email communication regarding this order.

20. * Please rate the timeliness of email communication regarding this order.

21. * Please rate the accuracy compared to your idea of an ideal purchase experience?

22. * Please rate the satisfaction with your idea of an ideal purchase experience?

23. * Please rate your satisfaction with the entire purchase experience?

24. * Please rate the purchase experience with your organization?

25. * Please rate the purchase experience with your organization?

26. * Please rate the purchase experience with your organization?

27. * Please rate the purchase experience with your organization?

28. * Please rate the purchase experience with your organization?

29. * Please rate the purchase experience with your organization?

30. * Please rate the purchase experience with your organization?

31. * Please rate the purchase experience with your organization?

32. * Please rate the purchase experience with your organization?

33. * Please rate the purchase experience with your organization?

34. * Please rate the purchase experience with your organization?

35. * Please rate the purchase experience with your organization?

36. * Please rate the purchase experience with your organization?

37. * Please rate the purchase experience with your organization?

38. * Please rate the purchase experience with your organization?

39. * Please rate the purchase experience with your organization?

40. * Please rate the purchase experience with your organization?

41. * Please rate the purchase experience with your organization?

42. * Please rate the purchase experience with your organization?

43. * Please rate the purchase experience with your organization?

44. * Please rate the purchase experience with your organization?

45. * Please rate the purchase experience with your organization?

46. * Please rate the purchase experience with your organization?

47. * Please rate the purchase experience with your organization?

48. * Please rate the purchase experience with your organization?

49. * Please rate the purchase experience with your organization?

50. * Please rate the purchase experience with your organization?

51. * Please rate the purchase experience with your organization?

52. * Please rate the purchase experience with your organization?

53. * Please rate the purchase experience with your organization?

54. * Please rate the purchase experience with your organization?

55. * Please rate the purchase experience with your organization?

56. * Please rate the purchase experience with your organization?

57. * Please rate the purchase experience with your organization?

58. * Please rate the purchase experience with your organization?

59. * Please rate the purchase experience with your organization?

60. * Please rate the purchase experience with your organization?

61. * Please rate the purchase experience with your organization?

62. * Please rate the purchase experience with your organization?

63. * Please rate the purchase experience with your organization?

64. * Please rate the purchase experience with your organization?

65. * Please rate the purchase experience with your organization?

66. * Please rate the purchase experience with your organization?

67. * Please rate the purchase experience with your organization?

68. * Please rate the purchase experience with your organization?

69. * Please rate the purchase experience with your organization?

70. * Please rate the purchase experience with your organization?

71. * Please rate the purchase experience with your organization?

72. * Please rate the purchase experience with your organization?

73. * Please rate the purchase experience with your organization?

74. * Please rate the purchase experience with your organization?

75. * Please rate the purchase experience with your organization?

76. * Please rate the purchase experience with your organization?

77. * Please rate the purchase experience with your organization?

78. * Please rate the purchase experience with your organization?

79. * Please rate the purchase experience with your organization?

80. * Please rate the purchase experience with your organization?

81. * Please rate the purchase experience with your organization?

82. * Please rate the purchase experience with your organization?

83. * Please rate the purchase experience with your organization?

84. * Please rate the purchase experience with your organization?

85. * Please rate the purchase experience with your organization?

86. * Please rate the purchase experience with your organization?

87. * Please rate the purchase experience with your organization?

88. * Please rate the purchase experience with your organization?

89. * Please rate the purchase experience with your organization?

90. * Please rate the purchase experience with your organization?

91. * Please rate the purchase experience with your organization?

92. * Please rate the purchase experience with your organization?

93. * Please rate the purchase experience with your organization?

94. * Please rate the purchase experience with your organization?

95. * Please rate the purchase experience with your organization?

96. * Please rate the purchase experience with your organization?

97. * Please rate the purchase experience with your organization?

98. * Please rate the purchase experience with your organization?

99. * Please rate the purchase experience with your organization?

100. * Please rate the purchase experience with your organization?

(Real BestBuy.com Survey)

Some Random Responses

"It is just garbage collect for Pete's sake!"

"i dont know"

"Person(s) who answered my calls were not very friendly. Very cold."

"it can't be."

Imagine analyzing this x6000

They don't even read it!

“Many shared stories about having huge datasets of qualitative responses but not touching it **because it's too wild to clean and interpret manually**”

“Some said **it's their single biggest painpoint and overhead** and **they walk away from projects or leave the data untouched**”

2 Part Solution:
A Better Survey +
Machine Learning Analysis

New Survey

Satisfaction Survey

What did you think?

Your answer

Rate your overall satisfaction

1

2

3

4

5



SUBMIT

Never submit passwords through Google Forms.

Comparison



Which would you rather do?

Satisfaction Survey

What did you think?

Your answer

Rate your overall satisfaction

1 2 3 4 5

View additional surveys through Design Forms

Why no simple surveys today?

Problem: We need to pay someone to read each and every response, and summarize the 'big' picture:

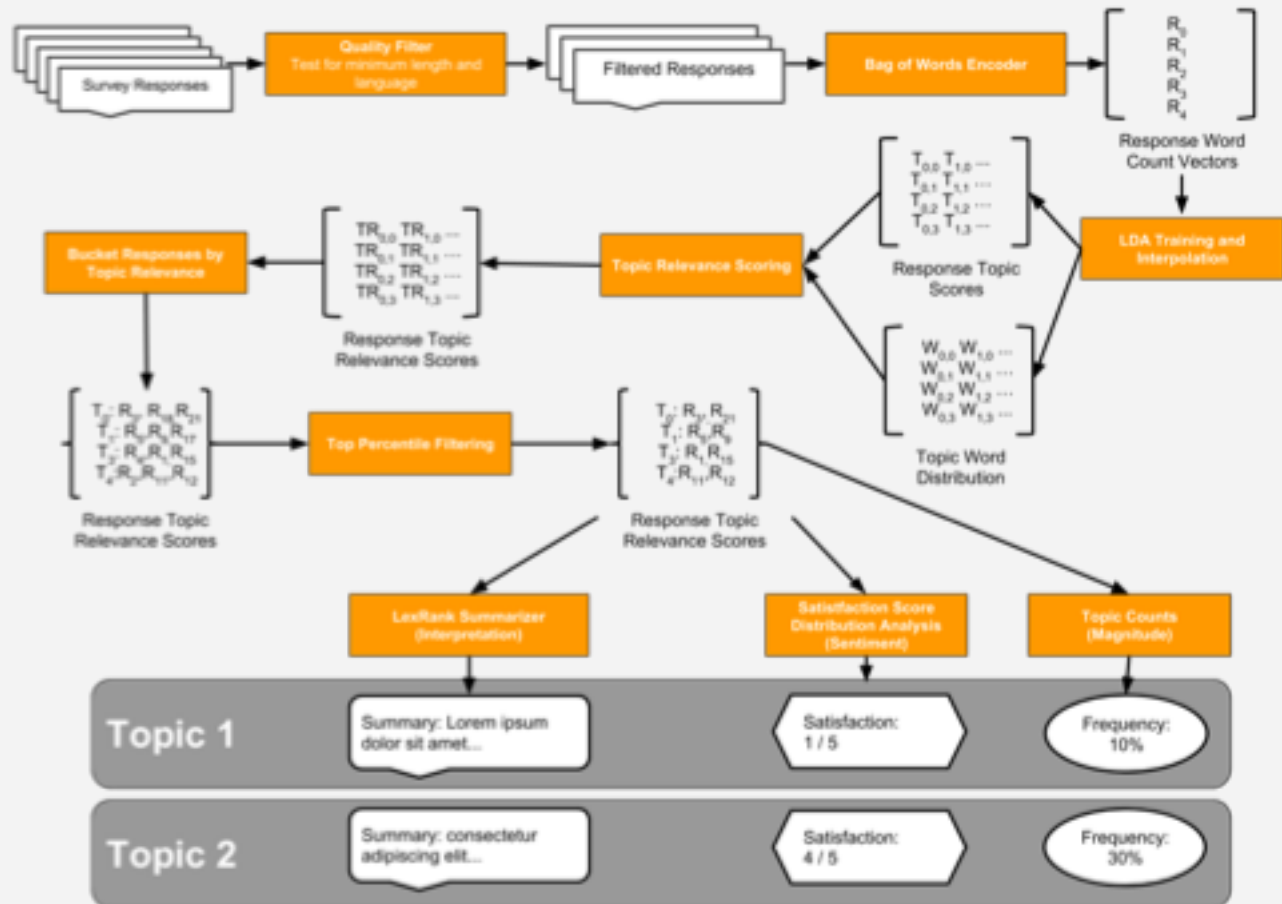
1. Difficult
2. Expensive
3. Vulnerable to bias

Solution: Machine Learning saves you from your labour intensive problems!

Combining a few ideas

1. **Topic Modelling** with Latent Dirichlet allocation.
 - a. But; topics won't be interpretable...
2. With **Automatic Summarization**...
 - a. But; sentiment and magnitude are not understood...
3. With satisfaction score **Statistical Analysis**
 - a. :)

The Big Picture



Part 2: Preprocess the Data

Bag of Words

“

There should be more
staff at the service
counter.

”

Bag of Words

“

There should be more
staff at the service
counter.

”



[
at
be
counter
more
service
should
staff
the
there
]

Weaknesses (1)

“

Toy dog

Dog toy

”



Weaknesses (2)

“

Buy old cars

=

Purchase used
automobiles

”



[buy
cars
old]

⊥

[automobiles
purchase
used]

In practice

```
bow = document
```

```
.lowercase()  
.split([' '])  
.split([punctuation_characters])  
.words_to_ids()
```



Word normalization (1)

walk

walked

walks

walking

Word normalization (1)

walk
walked
walks
walking



walk



Use **stemming** for fast word normalization

Word normalization (2)

“He found his true **calling** in life”

“He is **calling** her on the phone”

Word normalization (2)

“He found his true **calling** in life”

[Noun]

“He is **calling** her on the phone”

[Verb]



Use **lemmatization** for context-sensitive normalization

Language detection

“The service was very fast”

“Le service était très rapide”

Language detection

“The service was very fast”



“Le service était très rapide”



Filter if $P(\text{🇬🇧} \mid \text{text}) < 0.9$, just use BoW model

In practice

```
bow = document  
    .lowercase()  
    .split([' '])  
    .split([punctuation_characters])  
    .words_to_ids()
```

In practice

```
bow = document
```

```
.filter(is_english)  
.lowercase()  
.split([' '])  
.split([punctuation_characters])  
.lemmatize()  
.words_to_ids()
```


Part 3: Machine Learning & Analysis

Introducing LDA

Wikipedia: Latent Dirichlet Allocation (LDA) is a **generative** statistical model that allows sets of **observations** to be explained by **unobserved** groups that explain why some parts of the data are **similar**.

Introducing LDA

Wikipedia: Latent Dirichlet Allocation (LDA) is a **generative** statistical model that allows sets of **observations** to be explained by **unobserved** groups that explain why some parts of the data are **similar**.

Generative: Data and labels distribution is jointly learned $P(x,y)$ instead of discriminative $P(y|x)$

Observations: Using a training corpus

Unobserved: Without labels (unsupervised)

Similar: In order to cluster similar topics

LDA High Level

Input: List of free text documents (Bag of Words encoded), without labels.

Method: Training over n generations

Output: Scores for each topic per document

LDA By Analogy

Words

Words

Documents

Books

Topics

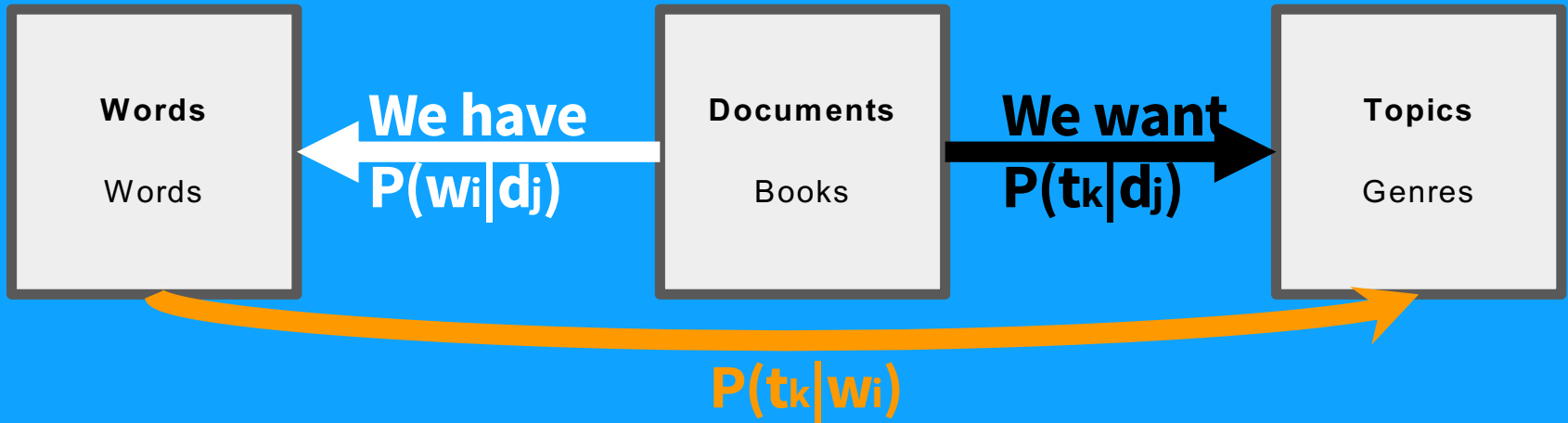
Genres

LDA By Analogy



LDA By Analogy

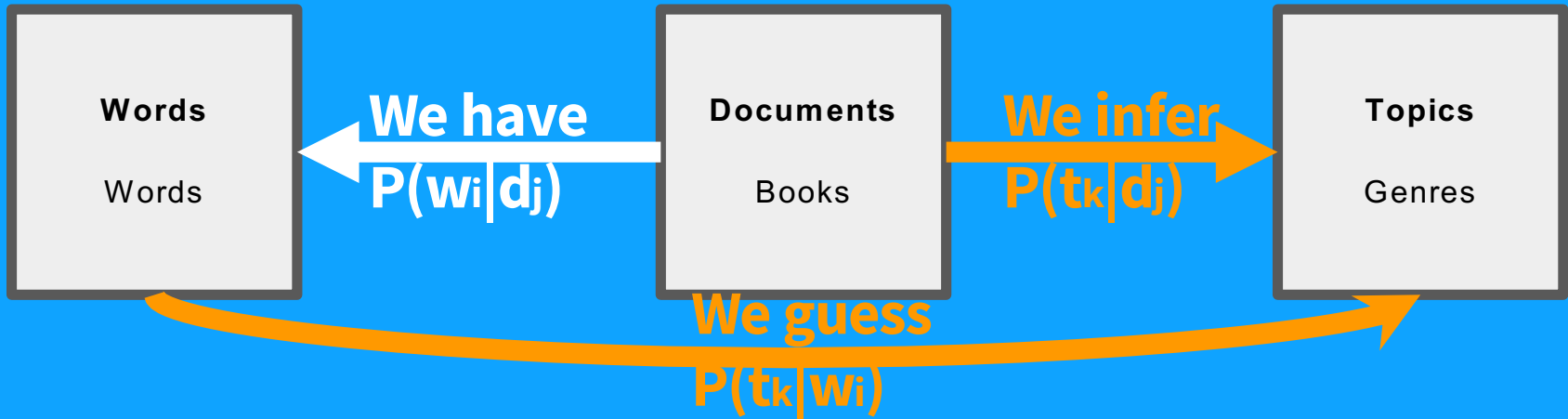
1. Init: Randomly give every word a topic - guess $P(t_k|w_i)$



LDA By Analogy

1. Init: Randomly give every word a topic - guess $P(t_k|w_i)$

$$P(t_k|d_j) = P(t_k|w_i) * P(w_i|d_j)$$



LDA By Analogy

- 1. Init:** Randomly give every word a topic - guess $P(t_k|w_i)$
- 2. Train:** For a given word w_i assign a new topic t_k by throwing a weighted die with probability $P(w_i \in t_k) = P(t_k|d_j) * P(w_i|t_k)$



LDA Problems!

Noise!

Every sentence gets a list of topic scores that sum to 1.

Response	“Ok, whatever this survey was lame”	“We need more highly trained staff in hospitals”
Topic 4 (Healthcare) LDA	75%	70%
Topic 7 (Transport) LDA	15%	10%
Topic 10 (Staff) LDA	10%	20%

LDA: Rank Relevance

New Idea: Score every response with a “relevance score”. It’s the normalized sum of the square probabilities of the words in the topic-word distribution.

$$\text{Rel}(\text{doc}, \text{topic}) = (\sum_{\text{word} \in \text{topic}} P(\text{word}|\text{topic})^2) / (\text{wordcount})$$

LDA: Rank Relevance

LDA Score vs LDA Rank Relevance Score

Response	“Ok, whatever this survey was lame”	“We need more highly trained staff in hospitals”
Topic 4 (Healthcare) LDA	75%	70%
Topic 4 Rank Relevance	0.00003	0.2
Topic 7 (Transport) LDA	15%	10%
Topic 7 Rank Relevance	0.000001	0.000002
Topic 10 (Staff) LDA	10%	20%
Topic 10 Rank Relevance	0.00000008	0.01

LDA: Rank Relevance Quality

Less good data is better than more noisy data.

Filtering Method:

1. Sort documents for each topic by rank relevance score.
2. Take top X% (dependent on data set ~25%-50%)

Document Summarization

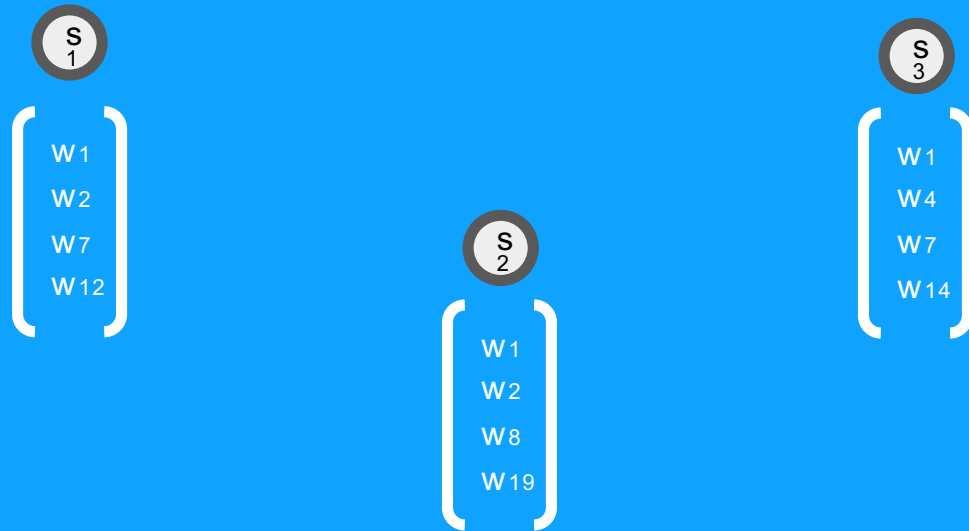
Goals:

- Short
- Most essential ideas
- Natural language

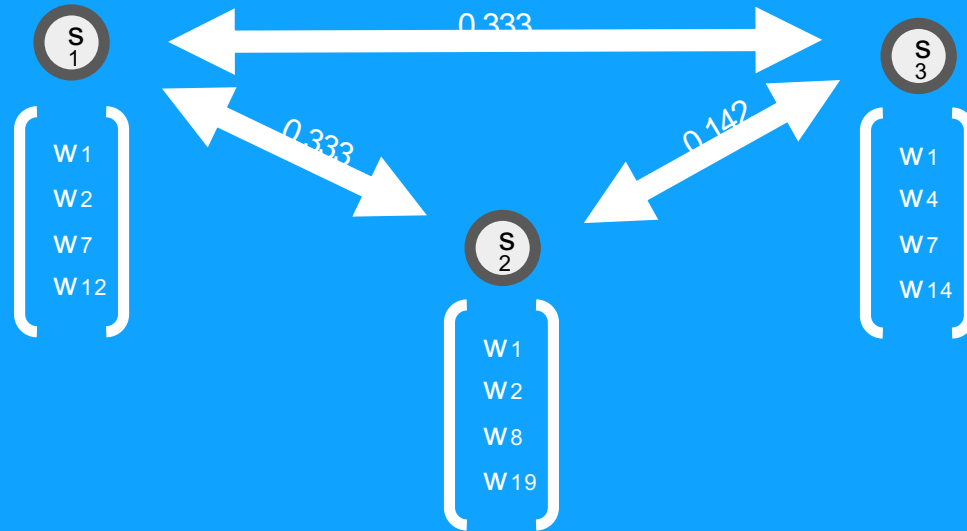
LexRank



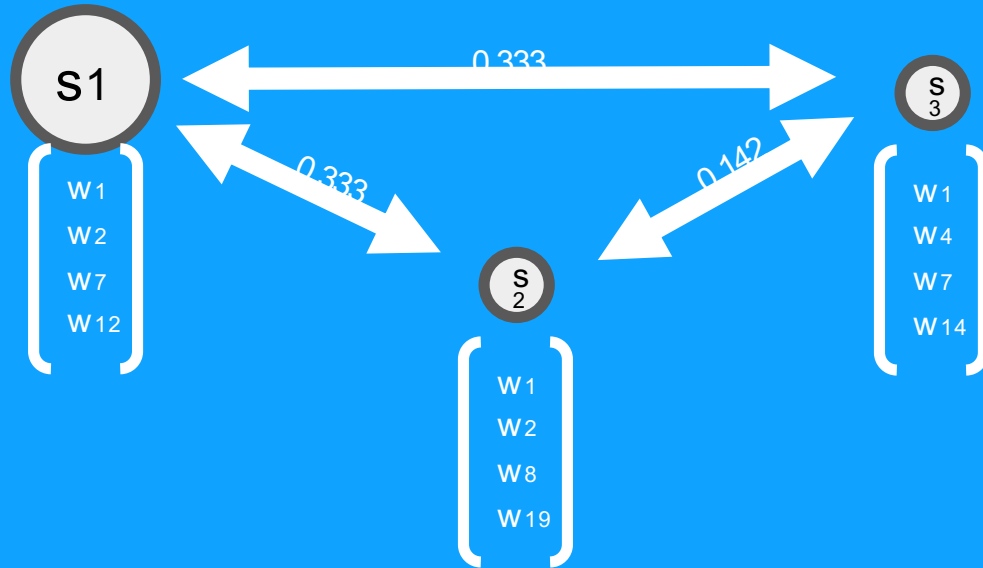
Sentences = Bag of Words



Compute Strength of Interaction



Node Centrality - PageRank



Document Summarization

Goals:

- Short
- Most essential ideas
- Natural language

top-k sentences

Only

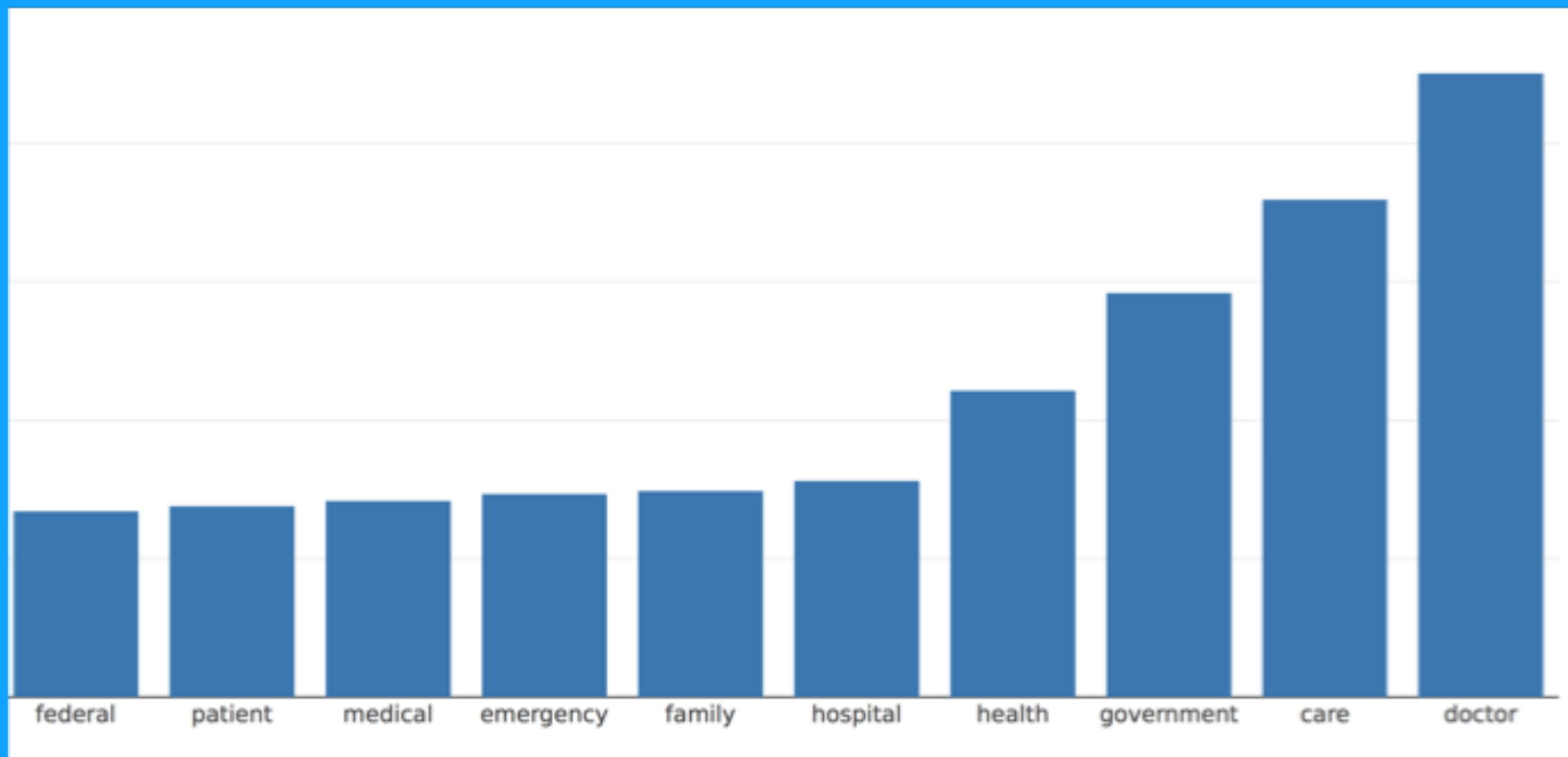
Node centrality

Full sentences

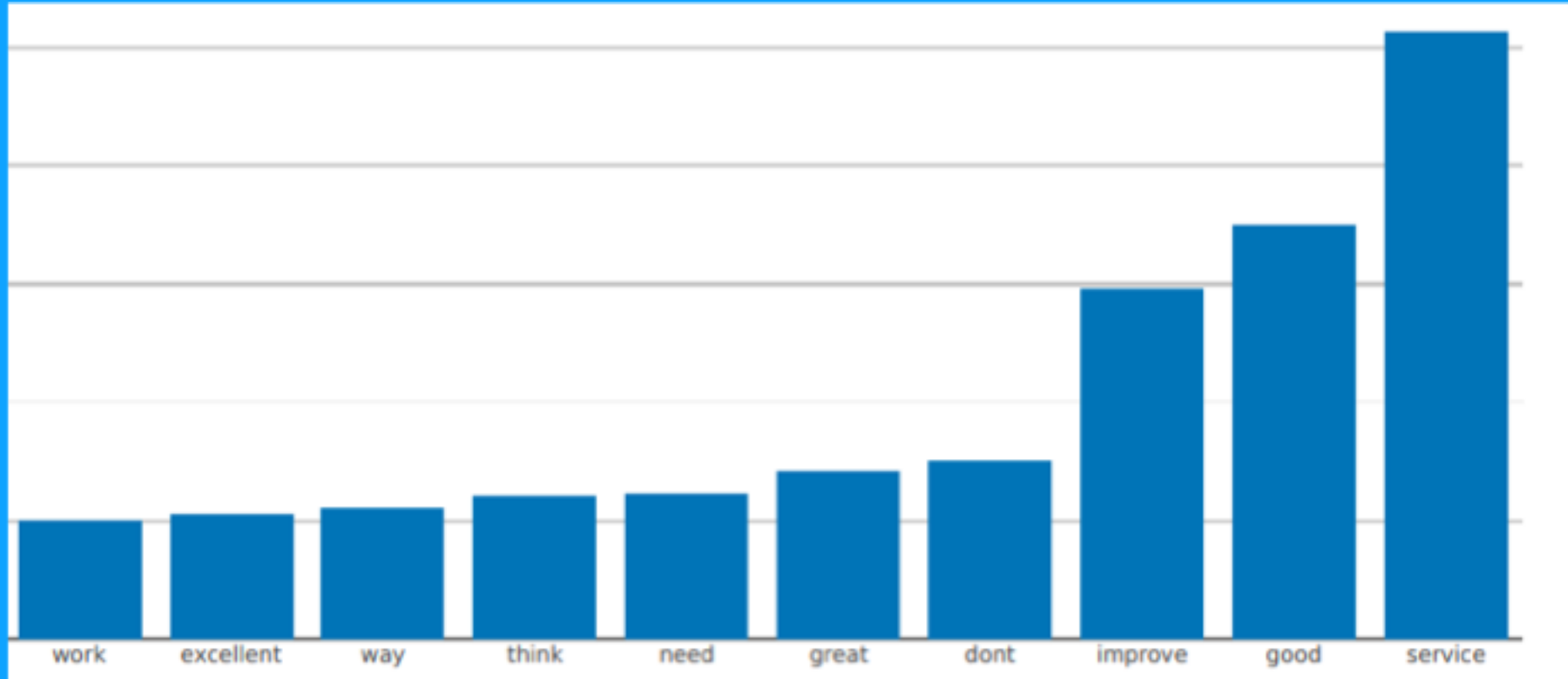
Part 4: Results

Example Topics discovered by LDA

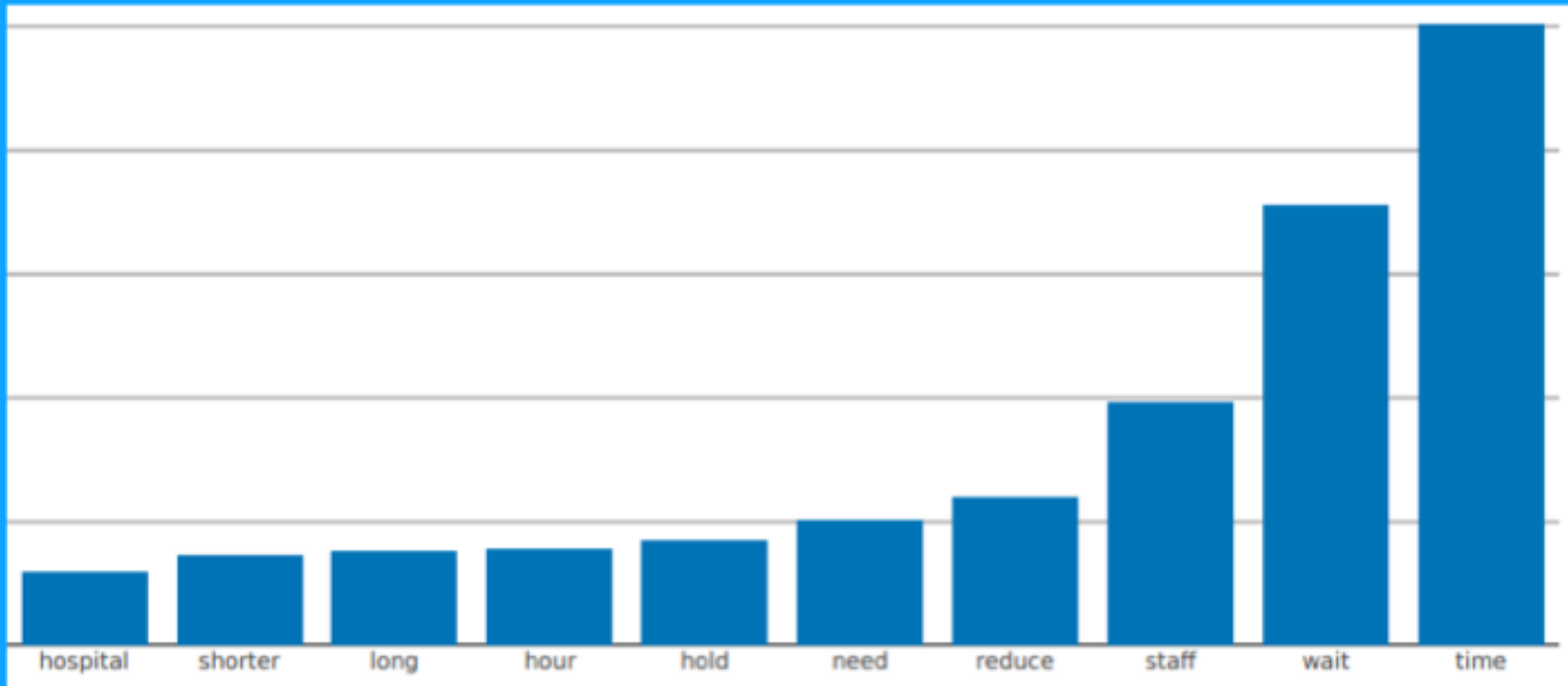
Topic 4



Topic 6



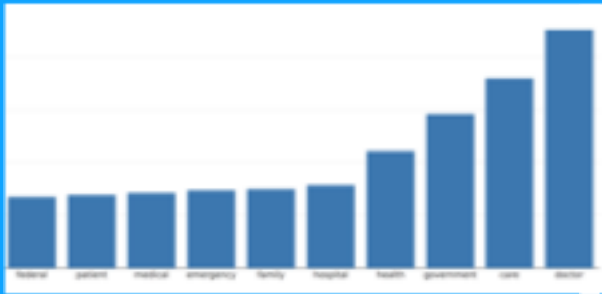
Topic 8



Summarization of Topics

Example Topics - Human Labels

Topic 4



Medical Services

Topic 6



Examples of good services

Topic 8



Waiting time

Example Topics - Summarization

Topic 4: Medical Services

WE NEED MORE DOCTORS,
NURSES, AND HOSPITAL
FACILITIES. How to improve
provincial government needs to
get/pay more doctors in the
province, they deserve it.
Provide more after hours clinics
for working families.

Topic 6: Examples of good services

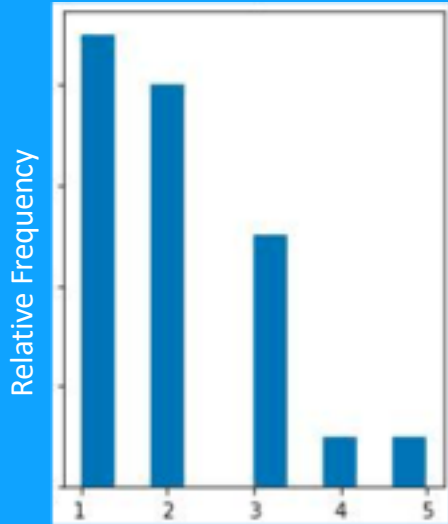
I do not know how to improve
the service, unless they have
more people working in the
office. It was a good service
and did not need any
improvement. I don't there is
anything currently that could
help this service to be
improved. I thought that the
service I got was very good. A
few more people on staff, but
overall service was good.

Topic 8: Waiting time

They take too much time
to come. Have staff
lunches/breaks at other
times. Easier contact info,
shorter wait times on
phones, being called back
on time. Next time I use
this service, I will
definitely use the internet
option as I don't have
time to wait in long lines.

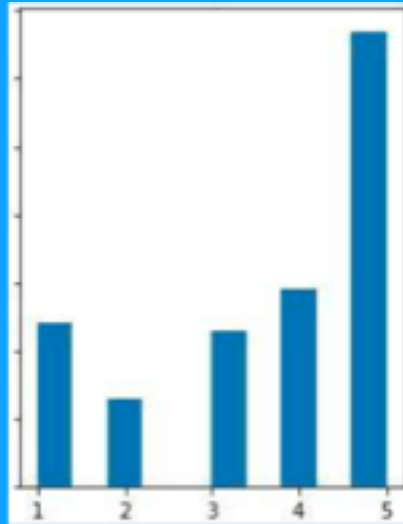
Statistical Analysis of Satisfaction

Topic 4



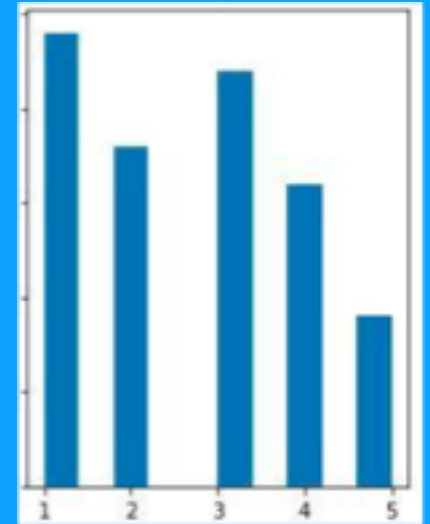
Medical Services

Topic 6



Examples of good services

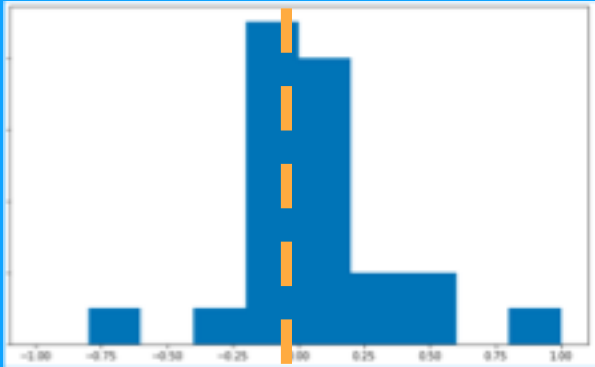
Topic 8



Waiting time

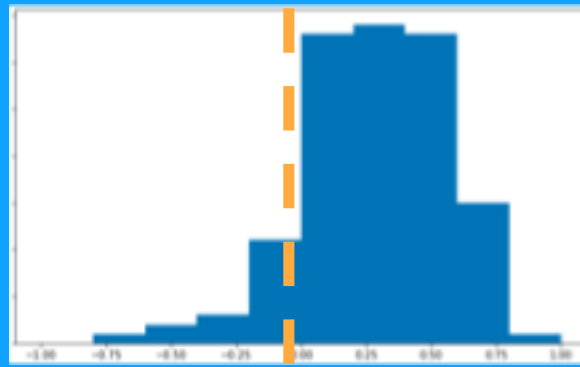
Sentiment Analysis

Topic 4



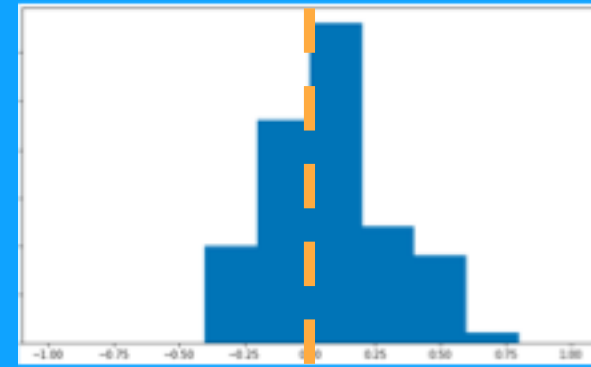
Medical Services

Topic 6



Examples of good services

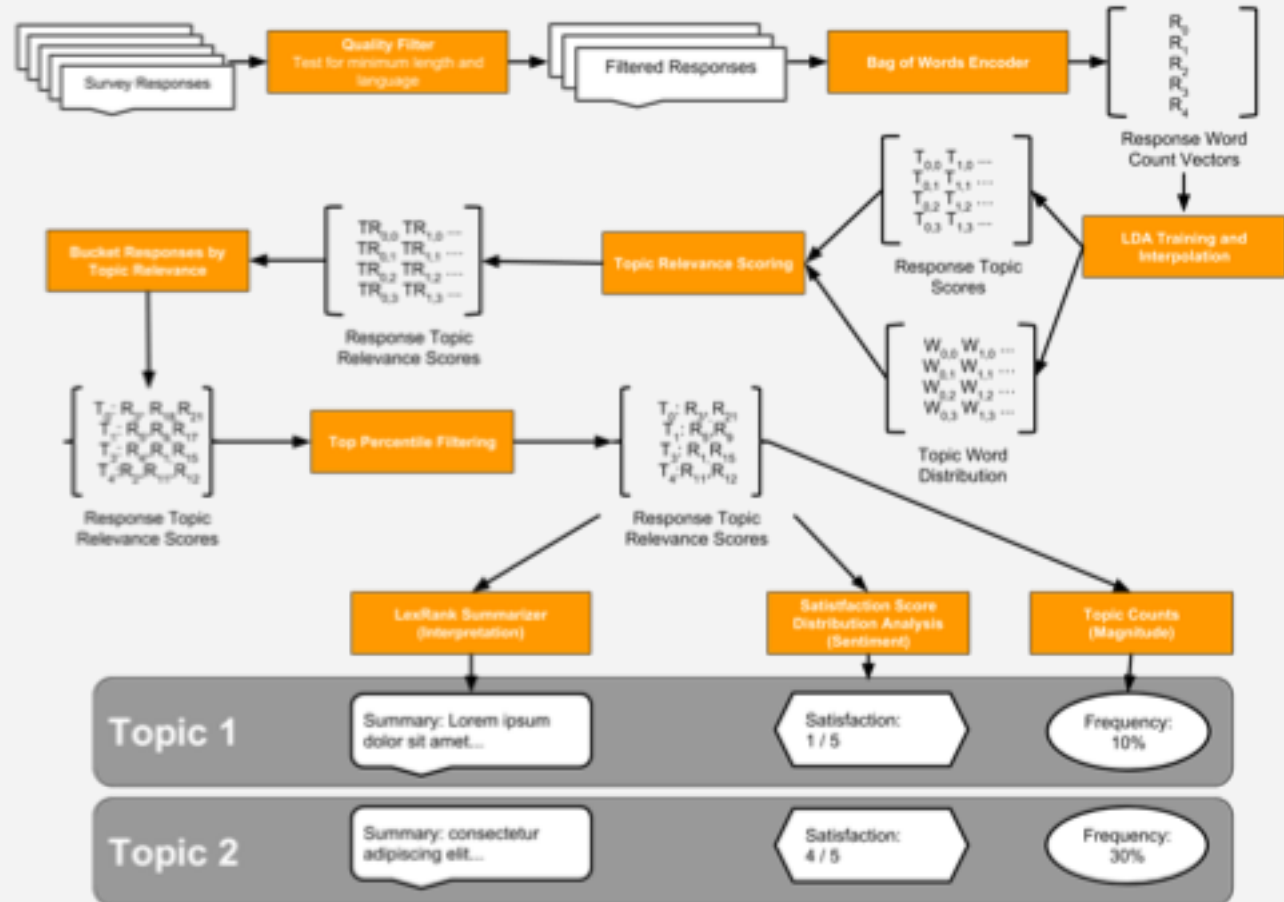
Topic 8



Waiting time

Recap

Using This



To This

Satisfaction Survey

What did you think?

Your answer

Rate your overall satisfaction

1

2

3

4

5



SUBMIT

Never submit passwords through Google Forms.

Pros:

- Automatic analysis
 - Cheaper
 - Less Bias
 - Survey less frustrating
- Not all responses equal
 - More writing =
More impact

Cons:

- Niche topics can be missed

Questions?

(P.S. We are both hiring!)



George Brova

CTO at Migacore Technologies

george@brova.com



Ling Zhang

Software Engineer at Aiden.ai

lz@ling.nz