

# fuys Team at the NTCIR-17 QA Lab-PoliInfo-4 Task

fuys team

Daigo Nishihara, Hokuto Ototake, Kenji Yoshimura

Fukuoka University

# About MBLink

- Tasks that link the mayor's statement to a table that deals with the budget.
- The mayor's statement and the table dealing with the budget have been changed to an **HTML file**.
  - The mayor's statement can be treated as **<p> tag**.
  - Tables dealing with budget can be treated as **<table> tag**.

それでは、平成29年度の予算編成についてですが、収入状況は、市税の伸びが期待できないことに加え、地方譲与税や交付金、さらには実質的な地方交付税の減少が見込まれ、引き続き大変厳しい状況にあります。

Mayor's statement

科目	本年度		前年度		比較	伸び率
	予算額	構成比	当初予算額	構成比		
	千円	%	千円	%	千円	%
1市税	13,188,100	23.9	13,202,700	23.3	△14,600	△0.1
2地方譲与税	321,001	0.6	332,001	0.6	△11,000	△3.3
3利子割交付金	10,000	0.0	20,000	0.0	△10,000	△50.0
4配当割交付金	37,000	0.1	54,000	0.1	△17,000	△31.5
5株式等譲渡所得割交付金	17,000	0.0	33,000	0.1	△16,000	△48.5
6地方消費税交付金	2,353,000	4.2	2,450,000	4.3	△97,000	△4.0
7ゴルフ場利用税交付金	35,000	0.1	36,000	0.1	△1,000	△2.8
8自動車取得税交付金	44,000	0.1	50,000	0.1	△6,000	△12.0
9国有提供施設等所在	400	0.0	400	0.0	—	—
9市町村助成交付金						
10地方特例交付金	35,800	0.1	35,000	0.1	800	2.3
11地方交付税	15,709,000	28.4	16,581,000	29.3	△872,000	△5.3
12交通安全対策特別交付金	15,000	0.0	16,000	0.0	△1,000	△6.3
13分担金及び負担金	281,881	0.5	299,804	0.5	△17,743	△5.9
14使用料及び手数料	981,333	1.8	1,010,317	1.8	△28,984	△2.9
15国庫支出金	11,558,579	20.9	11,544,974	20.4	11,805	0.1
16道支出金	3,156,249	5.7	3,189,853	5.8	△33,604	△1.1
17財産収入	52,872	0.1	55,137	0.1	△2,265	△4.1
18寄附金	2	0.0	1	0.0	1	100.0
19繰入金	720,387	1.3	287,812	0.5	432,775	150.5
20繰越金	1	0.0	1	0.0	—	—
21諸収入	2,790,248	5.0	2,803,025	4.9	△12,777	△0.5
22市債	3,986,600	7.2	4,838,900	8.2	△872,300	△14.5
歳入合計	55,271,433	100.0	56,639,525	100.0	△1,368,092	△2.4

# Proposed Method

- Finding related tables by focusing on cells.
- Cells with only monetary amounts, unit notations, or symbols are not used.
- We'll use the <p>tag text with the ID and the table cell text.
- **Binary classification task** with the two encoded and combined as input sentences.

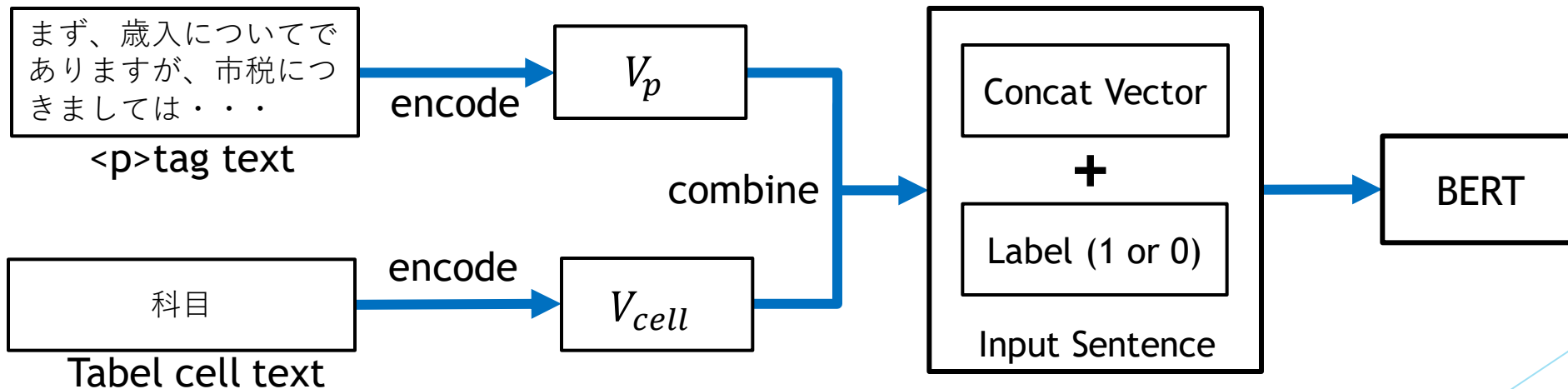


Table cells are extracted in **three different methods**.

# Proposed Method (Method 1)

- <p>tag and cell-by-cell learning and inference.
- Extract words from all table cells.
- Learning is performed with
  - **1** if the word exists in the related table.
  - **0** if the word does not exist in the related table.

科目	本年度		前年度	
	予算額	構成比	当初予算額	構成比
	千円	%	千円	%
1港湾整備事業費	244,463	40.7	240,558	41.2
2公債費	355,634	59.3	344,017	58.8
3予備費	100	0.0	100	0.0
歳出合計	600,197	100.0	584,675	100.0

extract cells

cell word list

```
[  
  “科目”  
  “本年度”  
  “前年度”  
  “比較”  
  ”伸び率”  
  “予算額”  
  “構成比”  
  . . .  
]
```

# Proposed Method (Method 1)

## Inference

- Look at each cell in the table one by one.
- If any one of the inference results outputs a related (**1**), then it is a related table.

まず、歳入についてありますが、市税につきましては・・・

<p> tag Text

Table1:	
	results
“科目”	0
“本年度”	0
“前年度”	<b>1</b>
“比較”	
...	

Table2

**Related Table**

Cell Word list

# Proposed Method (Method 2)

- Output **frequently appearing words** (e.g., “合計(total)”, “科目(subject)”) as related in Method 1.
  - Many unrelated tables are linked together.
- **Delete words that appear frequently.**
- Delete words that appear 6 or more times and redo learning and inference.

```
[  
"産業廃棄物等処分事業": 2  
"簡易水道事業": 2  
"合計": 36  
"科目": 18  
"1市税": 3  
"2地方譲与税": 3  
...  
]
```

Cell word list(text : number)

Delete words that  
appear more than 6

```
[  
"産業廃棄物等処分事業": 2  
"簡易水道事業": 2  
"1市税": 3  
"2地方譲与税": 3  
...  
]
```

New cell word list

# Proposed Method (Method 3)

- **<p>Tag and column-by-column** learning and inference.
- Extract words by table column.
- Join the extracted words with delimiters to form a single sentence.
- Use **[SEP]** and **“(comma)”** as delimiters.

科目	本年度		前年度	
	予算額	構成比	当初予算額	構成比
	千円	%	千円	%
1港湾整備事業費	244,463	40.7	240,558	41.2
2公債費	355,634	59.3	344,017	58.8
3予備費	100	0.0	100	0.0
歳出合計	600,197	100.0	584,675	100.0

Connect cell words by column with “,(comma)”

“科目,1港湾整備事業, . . . ,歳出合計”  
“本年度,予算額”  
“本年度,構成比”  
“前年度,当初予算額”  
“前年度,構成比”  
. . .

# Proposed Method (Method 3)

- Label each column and combine with <p>tags for learning.
- Label related tables with **1** and unrelated tables with **0**.
- If any one of the inference results outputs a related (**1**), then it is a related table.

まず、歳入についてで  
ありますが、市税につ  
きましては・・・

<p> tag Text

Table1:

[

“科目,1港湾整備事業,・・・,歳出合計”

“本年度,予算額”

“本年度,構成比”

“前年度,当初予算額”

“前年度,構成比”

・・・

]

Table2

Cell word list



# Proposed Method (Keyword extraction)

- There are multiple statements with the same meaning, but with different expressions and letterforms.
- **Keyword extraction** using TFIDF to eliminate fluctuations in statements.
- Use extracted words in place of <p>tag text.

また、手宮公園競技場への公認陸上競技場として必要な標識タイルの設置や、総合体育館への移動式バスケットゴール台などの施設整備により、スポーツ環境を整えてまいります。

2018 mayor's Statement

心身の健やかな成長を図るため、スポーツへの意欲を高める取り組みとしましては、市内小学校でプロスポーツ選手による出前授業を開催するほか、総合体育館への移動式バスケットゴール台整備や、手宮公園競技場への小学生用ハードルの整備により、競技環境の充実を図るとともに、大会の誘致に向けて取り組んでまいります。

2017 mayor's Statement

# Proposed Method (Keyword extraction)

- Extract the top 10 words
- Words are joined by delimiters to form a single sentence.
- Use **[SEP]** and “,(comma)” as delimiters.
- Use extracted words in place of <p>tag text in Method 3.

地方交付税につきましては、  
国の地方財政計画の伸び率な  
どを基本に、本市の特殊事情  
を勘案しながら積算した結果、  
臨時財政対策債を加えた実質  
的な地方交付税では、0.1%、  
1,900万円増の163億1,800万円  
を見込みました。

Mayor's statement

Keyword extraction  
by TFIDF

地方,税,交付,財政,円,積算,基本,  
伸び率,事情,勘案

Top 10 keywords

# Result

- The most accurate method is method 3, using ",(comma)" as the delimiter.

	Method 1	Method 2	Method 3 ([SEP])	Method 3 (,)
Result	1.43%	21.36%	27.23%	36.66%
Recall	88.79%	70.08%	31.28%	40.53%
Precision	00.74%	16.34%	46.75%	52.06%

	Keyword extraction ([SEP])	keyword extraction (,)
Result	11.63%	24.99%
Recall	12.55%	25.10%
Precision	49.58%	46.41%

# CONSIDERATION (Problem1)

- There is a <p>tag with no tables linked.
  - There were 22 out of 81 <p>tags in the output result that had nothing linked to them.
- The more rows in the table, the more words to join.
  - Words that provide material for human judgment (e.g., “一般会計(general accounting)”) are easily buried.

	会計別	本年度 予算額	前年度 当初予算額	比較	伸び率
		千円	千円	千円	%
	一般会計	56,639,525	55,047,861	1,591,664	2.9
	港湾整備事業	584,675	518,769	65,906	12.7
	青果物卸売市場事業	45,682	41,122	4,560	11.1
特	水産物卸売市場事業	35,119	37,164	△2,045	△5.5
	国民健康保険事業	18,407,404	18,472,977	△65,573	△0.4
別	住宅事業	819,995	902,375	△82,380	△9.1
	簡易水道事業	197,054	165,751	31,303	18.9
会	介護保険事業	14,915,813	14,382,147	533,666	3.7
	産業廃棄物処分事業	99,948	102,116	△2,168	△2.1
計	後期高齢者医療事業	1,971,630	2,051,509	△79,879	△3.9
	計	37,077,320	36,673,930	403,390	1.1
企	病院事業	11,714,578	10,928,488	786,090	7.2
業	水道事業	4,946,178	4,994,670	△48,492	△1.0
会	下水道事業	6,857,629	7,328,736	△471,107	△6.4
三上	産業廃棄物等処分事業	162,942	189,997	△27,055	△14.2
	計	23,681,327	23,441,891	239,436	1.0
合計		117,398,172	115,163,682	2,234,490	1.9

Tables not linked by our proposed method.  
In the Gold data, it is linked with many <p>tags.

# CONSIDERATION (Problem1)

## Solution

- Need to be flexible according to the shape of the table.
- Separate tables by half of a row if there are many rows
  - Text created by joining columns can be shortened.
  - It can prevent necessary words from being buried.

“會計別,一般會計,港灣整備事業,青  
果物卸売市場事業,水産物卸売市場  
事業,国民健康保険事業,住宅事業,  
簡易水道事業,介護保険事業,産業  
廃棄物処分事業,後期高齢者医療事  
業,病院事業,水道事業,下水道事業,  
産業廃棄物等処分事業,合計”

separate

“會計別,一般會計,港灣整備事業,青  
果物卸売市場事業,水産物卸売市場  
事業,国民健康保険事業,住宅事業,  
簡易水道事業”

“介護保険事業,産業廃棄物処分事  
業,後期高齢者医療事業,病院事業,  
水道事業,下水道事業,産業廃棄物  
等処分事業,合計”

# CONSIDERATION (Problem2)

- Distributed data may not be well maintained.
- The tables with lists are linked in the Gold data.
- The tables with individual details are not link in the Gold data.
- In our proposed method, the following table was linked to the statement about “地方交付税(local allocation tax)”.

交付税						
予算科目	本年度予算額	前年度予算額	比較	節		説明
				区分	金額	
	千円	千円	千円		千円	千円
<b>11 地方交付税</b>	<b>16,581,000</b>	<b>15,340,000</b>	<b>1,241,000</b>			
1地方交付税	16,581,000	15,340,000	1,241,000			
1地方交付税	16,581,000	15,340,000	1,241,000	1地方交付税	16,581,000	普通交付税15,615,000
						特別交付税966,000

Table with details of local allocation tax

**Thank you for listening.**