

予測モデルでのDMD活用例

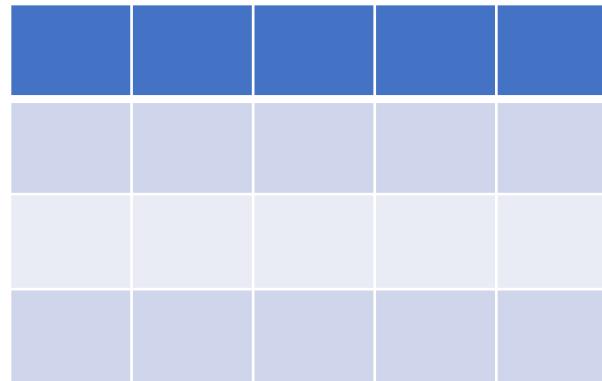
2019年3月20日

第三回勉強会

IPAデータ活用推進グループ 豊田 耕司

予測モデルとは？

原因データ x



$$y = f(x)$$



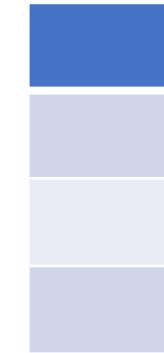
予測モデル f

例えば...

気温、湿度、天気

気温、湿度、天気、
ビールの売上の
過去の実績データ
で学習したモデル

目的データ y



ビールの売上

予測モデルが活用されている場面

- 発生確率を予測する
 - 例：国勢調査の傾向より、マンションの購買層を予測する
 - データの分類
 - 例：診療データより、疾病のリスク群を分類する
 - データの相関性を見つける
 - 例：匿名診療データより、血圧と病気の相関性を見つける
 - 例：地域コミュニティーごとの避難所種別の過不足傾向を見つける
- などなど...

予測モデルを考える

- 対象のデータ項目の意味より、予測モデルのデータモデルを考える

標高	電話番号	内線番号	市区町村コード	都道府県名	市区町村名	災害種別_洪水	崖崩れ、土石流及	災害種別_高潮	災害種別_地震	災害種別_津波	害種別_大規模な災害種別_内水氾濫	災害種別_火山噴出	指定避難所との重複	想定収容人数	象となる町会・自治会	URL	備考
	0797-71-8886		282146	兵庫県	宝塚市	1			1		1			1			
	0797-74-3448		282146	兵庫県	宝塚市	1	1		1		1			1			
	0797-71-0492		282146	兵庫県	宝塚市	1	1		1		1			1			
	0797-71-8451		282146	兵庫県	宝塚市	1	1		1		1			1			
	0797-73-3305		282146	兵庫県	宝塚市	1		1		1				1			
	0797-72-6581		282146	兵庫県	宝塚市		1		1		1			1			
	0797-71-5511		282146	兵庫県	宝塚市	1	1		1		1			1			
	0797-71-5330		282146	兵庫県	宝塚市	1	1		1		1			1			
	0797-72-5586		282146	兵庫県	宝塚市	1	1		1		1			1			
	0797-73-7348		282146	兵庫県	宝塚市		1		1		1			1			
	0797-73-3297		282146	兵庫県	宝塚市	1	1		1		1			1			
	0798-51-1132		282146	兵庫県	宝塚市										1		
	0798-52-1166		282146	兵庫県	宝塚市										1		
	072-757-8810		282146	兵庫県	宝塚市										1		
	0797-89-0224		282146	兵庫県	宝塚市										1		
	0797-89-4145		282146	兵庫県	宝塚市										1		
	0797-87-5744		282146	兵庫県	宝塚市										1		
	0797-84-8997		282146	兵庫県	宝塚市										1		
	0797-87-0091		282146	兵庫県	宝塚市										1		
	0797-88-3137		282146	兵庫県	宝塚市										1		
	0797-87-0296		282146	兵庫県	宝塚市										1		
	0797-87-0019		282146	兵庫県	宝塚市										1		
	0797-87-0292		282146	兵庫県	宝塚市		1		1		1				1		
	0797-86-7771		282146	兵庫県	宝塚市	1			1		1				1		
	0797-87-0451		282146	兵庫県	宝塚市	1	1		1		1				1		
	0797-87-4405		282146	兵庫県	宝塚市	1	1		1		1				1		
	0797-84-2441		282146	兵庫県	宝塚市	1	1		1		1				1		
	0797-89-5412		282146	兵庫県	宝塚市	1			1		1				1		
	0797-88-7512		282146	兵庫県	宝塚市	1	1		1		1				1		
	0797-89-3010		282146	兵庫県	宝塚市	1		1	1		1				1		
	0797-88-6492		282146	兵庫県	宝塚市	1	1		1		1				1		
	0797-88-2031		282146	兵庫県	宝塚市	1	1		1		1				1		
	0797-88-5322		282146	兵庫県	宝塚市	1			1		1				1		
	0797-88-1201		282146	丘陵地	宝塚市	1	1		1		1				1		

表データを見る！

...難しい

予測モデルを考える

- 対象のデータ項目の意味より、予測モデルのデータモデルを考える

標高	電話番号	内線番号	市区町村コード	都道府県名	市区町村名	災害種別_洪水	災害種別_崖崩れ、土石流	災害種別_高潮	災害種別_地震	災害種別_津波	災害種別_大規模な火事	災害種別_内水氾濫	災害種別_火山現象	指定避難所との重複	想定収容人数	象となる町会・自治会	URL	備考
	0797-71-8886		282146	兵庫県	宝塚市	1												
	0797-74-3448		282146	兵庫県	宝塚市	1												
	0797-71-0492		282146	兵庫県	宝塚市	1												
	0797-71-8451		282146	兵庫県	宝塚市	1												
	0797-73-3305		282146	兵庫県	宝塚市	1												
	0797-73-6501		282146	兵庫県	宝塚市	1												

項目の意味といえば、
共通語彙基盤！

072-757-8810	282146	兵庫県	宝塚市	1
0797-89-0224	282146	兵庫県	宝塚市	
0797-89-4145	282146	兵庫県	宝塚市	1
0797-87-5744	282146	兵庫県	宝塚市	1
0797-84-8997	282146	兵庫県	宝塚市	1
0797-87-0091	282146	兵庫県	宝塚市	
0797-88-3137	282146	兵庫県	宝塚市	1
0797-87-0296	282146	兵庫県	宝塚市	1
0797-87-0019	282146	兵庫県	宝塚市	
0797-87-0292	282146	兵庫県	宝塚市	
0797-86-7771	282146	兵庫県	宝塚市	1
0797-87-0451	282146	兵庫県	宝塚市	1
0797-87-4405	282146	兵庫県	宝塚市	1
0797-84-2441	282146	兵庫県	宝塚市	1
0797-89-5412	282146	兵庫県	宝塚市	1
0797-88-7512	282146	兵庫県	宝塚市	1
0797-89-3010	282146	兵庫県	宝塚市	
0797-88-6492	282146	兵庫県	宝塚市	1
0797-88-2031	282146	兵庫県	宝塚市	1
0797-88-5322	282146	兵庫県	宝塚市	1
0797-88-1201	282146	兵庫県	宝塚市	1

datamodel.txt

```
use ic:施設型>ic:関与[都道府県名]>ic:関与者>ic:住所>ic:都道府県;
use ic:施設型>ic:関与[市区町村名]>ic:関与者>ic:住所>ic:市区町村;
use ic:施設型>ic:関与[市区町村コード]>ic:関与者>ic:ID>ic:識別値;
use ic:施設型>ic:種別コード[災害種別_洪水]>ic:識別値;
use ic:施設型>ic:種別コード[災害種別_地震]>ic:識別値;
use ic:施設型>ic:種別コード[災害種別_崖崩れ、土石流及び地滑り]>ic:識別値;
use ic:施設型>ic:種別コード[災害種別_高潮]>ic:識別値;
use ic:施設型>ic:種別コード[災害種別_津波]>ic:識別値;
use ic:施設型>ic:種別コード[災害種別_大規模な火事]>ic:識別値;
use ic:施設型>ic:種別コード[災害種別_内水氾濫]>ic:識別値;
use ic:施設型>ic:種別コード[災害種別_火山現象]>ic:識別値;
use ic:施設型>ic:種別コード[指定緊急避難場所]>ic:識別値;
use ic:施設型>ic:種別コード[指定避難所との重複]>ic:識別値;
use ic:施設型>ic:種別コード[福祉避難施設所]>ic:識別値;
use ic:施設型>ic:種別コード[その他の避難場所]>ic:識別値;
use ic:施設型>ic:収容人数>ic:数値;
use ic:施設型>ic:記述[搬入可能なトラックの大きさ]>ic:説明;
```

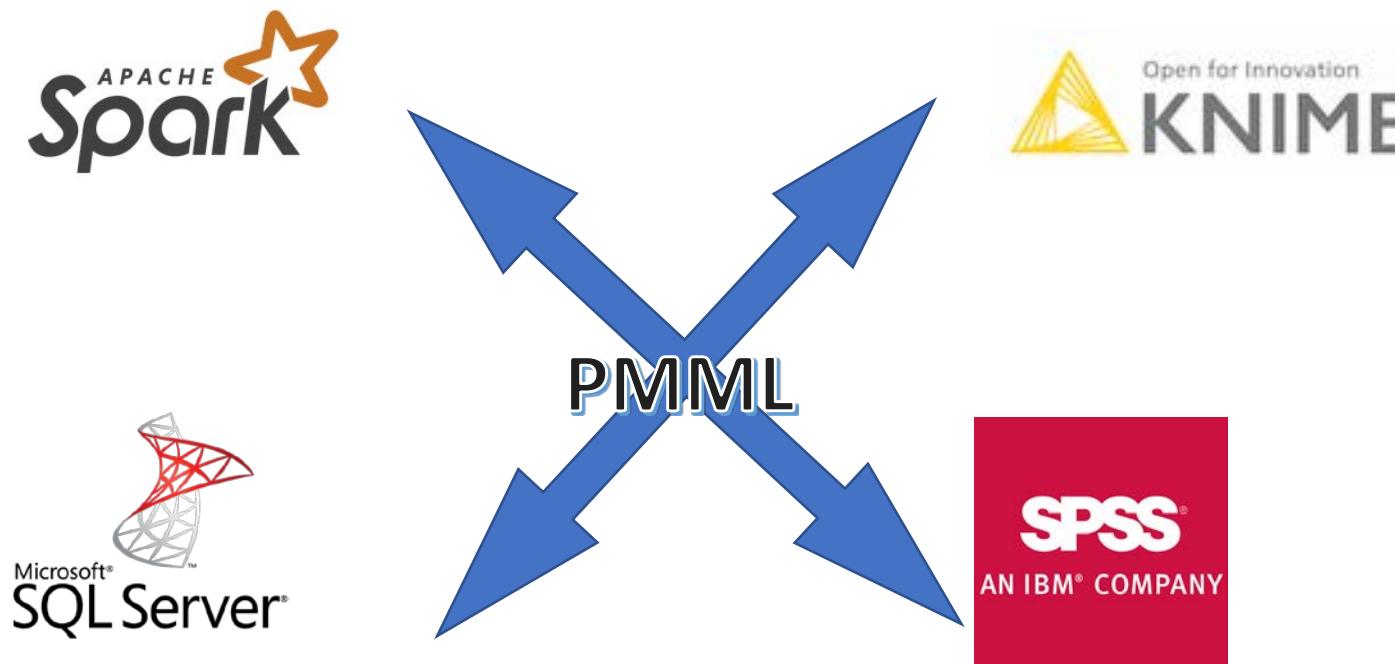
今日のデモ（予測モデルでの活用例）

1. DMDからPMMLを生成
2. KNIMEというPMMLに読ませる
3. KNIMEの画面からPMMLを見てみる
 1. 実データを見てみる
 2. DMDを見てみる
 3. PMMLに反映してみる

PMMIとは



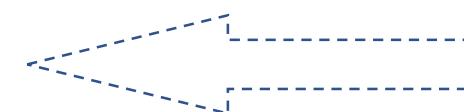
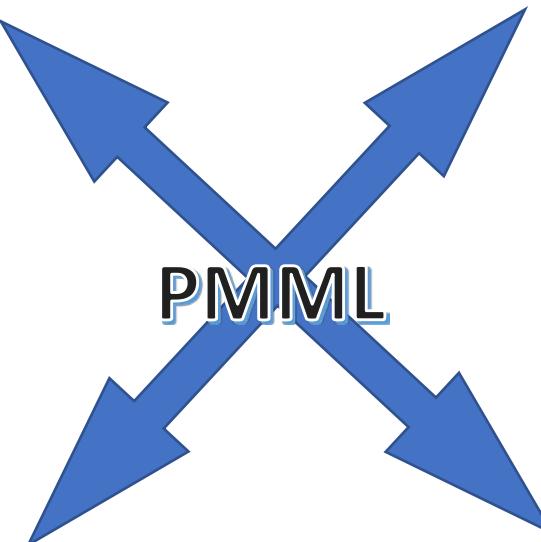
- 予測モデルマークアップランゲージ (Predictive Model Markup Language)
- 異なるデータマイニング製品間で予測・分析モデルの交換・共有を可能にするXMLの拡張



PMMIとは



- 予測モデルマークアップランゲージ (Predictive Model Markup Language)
- 異なるデータマイニング製品間で予測・分析モデルの交換・共有を可能にするXMLの拡張



IMIもアプローチ
できないかな??

1.DMDからPMMLを作成

```
-<PMML version="4.1">
  <Header copyright="example.com"/>
  -<DataDictionary numberOfFields="31">
    +<DataField dataType="string" name="ID" taxonomy="ic:施設型&gt;ic:ID&gt;ic:識別値" optype="categorical"></DataField>
    +<DataField dataType="string" name="名称" taxonomy="ic:施設型&gt;ic:名称&gt;ic:表記" optype="categorical"></DataField>
    +<DataField dataType="string" name="名称_カナ" taxonomy="ic:施設型&gt;ic:名称&gt;ic:カナ表記" optype="categorical"></DataField>
    +<DataField dataType="string" name="住所" taxonomy="ic:施設型&gt;ic:住所&gt;ic:表記" optype="categorical"></DataField>
    +<DataField dataType="string" name="方書" taxonomy="ic:施設型&gt;ic:住所&gt;ic:方書" optype="categorical"></DataField>
    +<DataField dataType="string" name="緯度" taxonomy="ic:施設型&gt;ic:地理座標&gt;ic:緯度" optype="categorical"></DataField>
    +<DataField dataType="double" name="標高" taxonomy="ic:施設型&gt;ic:地理座標&gt;ic:測地高度&gt;ic:数値" optype="continuous"></DataField>
    +<DataField dataType="string" name="経度" taxonomy="ic:施設型&gt;ic:地理座標&gt;ic:経度" optype="categorical"></DataField>
    +<DataField dataType="string" name="電話番号" taxonomy="ic:施設型&gt;ic:連絡先&gt;ic:電話番号" optype="categorical"></DataField>
    +<DataField dataType="string" name="内線番号" taxonomy="ic:施設型&gt;ic:連絡先&gt;ic:内線番号" optype="categorical"></DataField>
    +<DataField dataType="string" name="都道府県名" taxonomy="ic:施設型&gt;ic:関与[都道府県名][ic:役割=&apos;設置主体&apos;]&gt;ic:関与者" optype="categorical"></DataField>
    +<DataField dataType="string" name="市区町村名" taxonomy="ic:施設型&gt;ic:関与[市区町村名][ic:役割=&apos;設置主体&apos;]&gt;ic:関与者" optype="categorical"></DataField>
    +<DataField dataType="string" name="市区町村コード" taxonomy="ic:施設型&gt;ic:関与[市区町村コード][ic:役割=&apos;設置主体&apos;]" optype="categorical"></DataField>
    +<DataField dataType="string" name="災害種別_洪水" taxonomy="ic:施設型&gt;ic:種別コード[災害種別_洪水][ic:種別=&apos;設置主体&apos;]" optype="categorical"></DataField>
    +<DataField dataType="string" name="災害種別_地震" taxonomy="ic:施設型&gt;ic:種別コード[災害種別_地震][ic:種別=&apos;災害種別_地震&apos;]" optype="categorical"></DataField>
    +<DataField dataType="string" name="災害種別_崖崩れ、土石流及び地滑り" taxonomy="ic:施設型&gt;ic:種別コード[災害種別_崖崩れ、土石流及び地滑り]" optype="categorical"></DataField>
    +<DataField dataType="string" name="災害種別_高潮" taxonomy="ic:施設型&gt;ic:種別コード[災害種別_高潮][ic:種別=&apos;災害種別_高潮&apos;]" optype="categorical"></DataField>
    +<DataField dataType="string" name="災害種別_津波" taxonomy="ic:施設型&gt;ic:種別コード[災害種別_津波][ic:種別=&apos;災害種別_津波&apos;]" optype="categorical"></DataField>
  -<DataField dataType="string" name="災害種別_大規模な火事" taxonomy="ic:施設型&gt;ic:種別コード[災害種別_大規模な火事][ic:種別=&apos;災害種別_大規模な火事&apos;]" optype="categorical"></DataField>
    <Value value="自分で入れる"/>
```

参照 : 01_pmml(元データ).xml

2.KNIMEにPMMLを読み込ませる

KNIME Explorer

- EXAMPLES (knime-guest@http://public-server.knime.com:80/tor)
- LOCAL (Local Workspace)
 - Example Workflows
 - PMML_TEST_20190315

Workflow Coach

Node recommendations only available with usage data reporting. Click here to enable.

Node Repository

- IO
- Manipulation
- Views
- Analytics
- Database
- Other Data Types
- Structured Data
- Scripting
- Tools & Services
- Workflow Control
- Reporting

KNIME Analytics Platform - /Users/k-toyota/knime-workspace

KNIME Explorer

Node Description

Table Writer

PMML Reader

PMML To Cell

Node 1

Node 2

Node 3

Node 4

CSV Writer

Outline

Console

```

PMML Reader --> PMML To Cell
PMML To Cell --> Table Writer
PMML To Cell --> CSV Writer
  
```

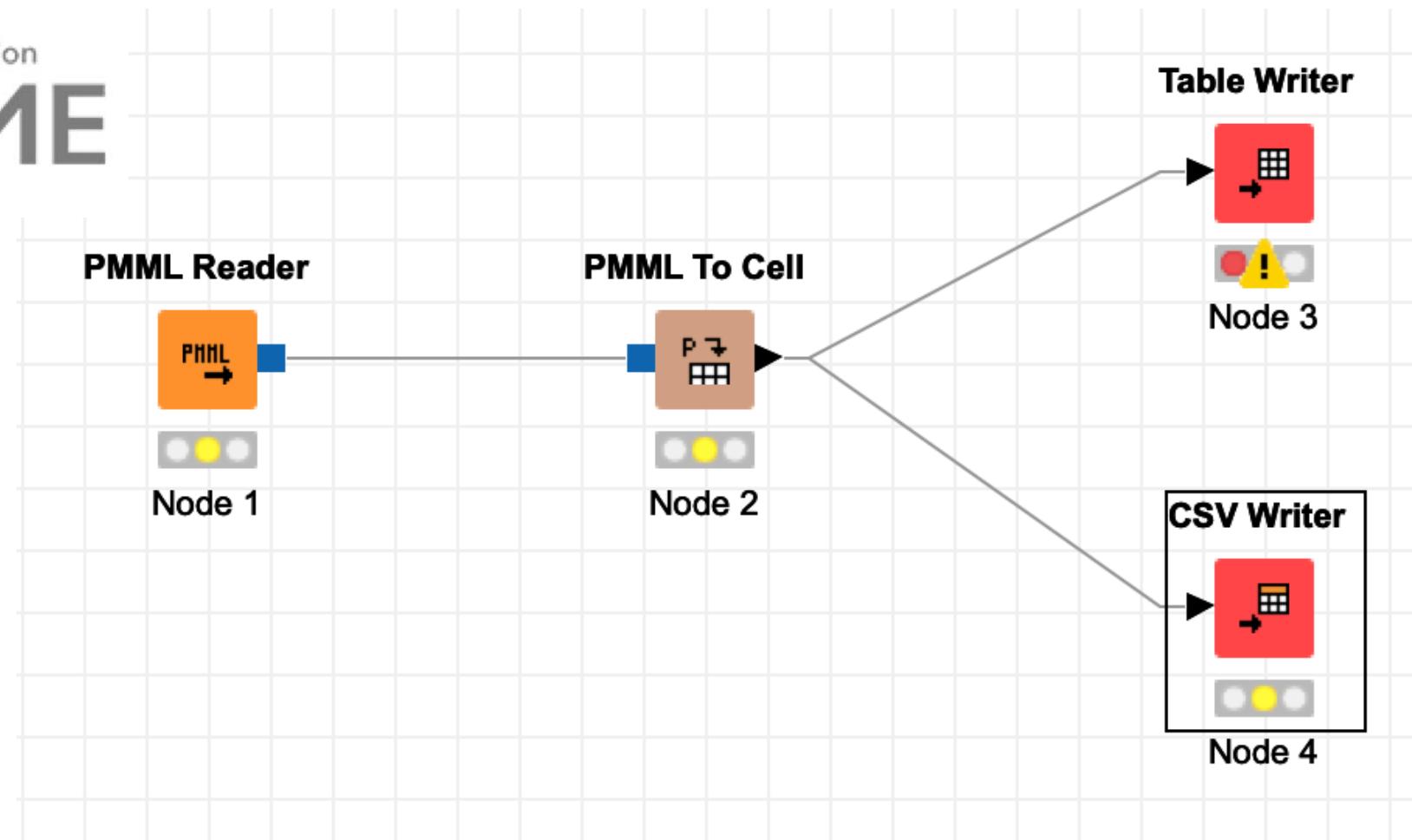
KNIME Console

```

*****
***      Welcome to KNIME Analytics Platform v3.7.1.v201901281154 ***
***      Copyright by KNIME AG, Zurich, Switzerland ***
*****
Log file is located at: /Users/k-toyota/knime-workspace/.metadata/knime/knime.log
WARN  Table Writer      0:3      Output file '/Users/k-toyota/00_Develop/02_imiライブラリ/04.
  
```

9

2 .KNIMEにPMMLを読み込ませる



3. KNIMEの画面からPMMLを見てみる

Columns: 31	Column Type	Column Index	Color Handler	Size Handler	Shape Handler	Filter Handler	Lower Bound	Upper Bound	Value 0
ID	String	0					?	?	自分で入れる
名称	String	1					?	?	自分で入れる
名称_カナ	String	2					?	?	自分で入れる
住所	String	3					?	?	自分で入れる
方書	String	4					?	?	自分で入れる
緯度	String	5					?	?	自分で入れる
標高	Number (double)	6					0	0	?
経度	String	7					?	?	自分で入れる
電話番号	String	8					?	?	自分で入れる
内線番号	String	9					?	?	自分で入れる
都道府県名	String	10					?	?	自分で入れる
市区町村名	String	11					?	?	自分で入れる
市区町村コード	String	12					?	?	自分で入れる
災害種別_洪水	String	13					?	?	自分で入れる
災害種別_地震	String	14					?	?	自分で入れる
災害種別_崖崩れ、土石流及び地滑り	String	15					?	?	自分で入れる
災害種別_高潮	String	16					?	?	自分で入れる
災害種別_津波	String	17					?	?	自分で入れる
災害種別_大規模な火事	String	18					?	?	自分で入れる
災害種別_内水氾濫	String	19					?	?	自分で入れる
災害種別_火山現象	String	20					?	?	自分で入れる
指定緊急避難場所	String	21					?	?	自分で入れる
指定避難所	String	22					?	?	自分で入れる
福祉避難所	String	23					?	?	自分で入れる
その他の避難場所	String	24					?	?	自分で入れる
想定収容人数	Number (double)	25					0	0	?
搬入可能なトラックの大きさ	String	26					?	?	自分で入れる
対象となる町会・自治会	String	27					?	?	自分で入れる
URL	String	28					?	?	自分で入れる
備考	String	29					?	?	自分で入れる
リンクID	String	30					?	?	自分で入れる

3. KNIMEの画面からPMMLを見てみる

Columns: 31	Column Type	Column Index	Color Handler	Size Handler	Shape Handler	Filter Handler	Lower Bound	Upper Bound	Value 0
ID	String	0					?	?	自分で入れる
名称	String	1					?	?	自分で入れる
名称_カナ	String	2					?	?	自分で入れる
住所	String	3					?	?	自分で入れる
方書	String	4					?	?	自分で入れる
緯度	String	5					?	?	自分で入れる
標高	Number (double)	6					0	0	?
経度	String	7							
電話番号	String	8							
内線番号	String	9							
都道府県名	String	10							
市区町村名	String	11							
市区町村コード	String	12							
災害種別_洪水	String	13							
災害種別_地震	String	14							
災害種別_崖崩れ、土石流及び地滑り	String	15							
災害種別_高潮	String	16							
災害種別_津波	String	17							
災害種別_大規模な火事	String	18							
災害種別_内水氾濫	String	19							
災害種別_火山現象	String	20							
指定緊急避難場所	String	21							
指定避難所	String	22							
福祉避難所	String	23							
その他の避難場所	String	24					?	?	自分で入れる
想定収容人数	Number (double)	25					0	0	?
搬入可能なトラックの大きさ	String	26					?	?	自分で入れる
対象となる町会・自治会	String	27					?	?	自分で入れる
URL	String	28					?	?	自分で入れる
備考	String	29					?	?	自分で入れる
リンクID	String	30					?	?	自分で入れる

例えば「災害種別」
 これらの項目から予測モデル
 は関連性を理解できるの?
 難しいかも...

データを正規化して考えてみる!

3. KNIMEの画面からPMMLを見てみる—実データを見てみる

N	O	P	Q	R	S	T	U
災害種別_洪水	崖崩れ、土石流及 ての	災害種別_高潮	災害種別_地震	災害種別_津波	皆種別_大規模な災害	災害種別_内水氾濫	災害種別_火山現象
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
	1		1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1		1		1		
1			1		1		
1	1		1		1		
1	1		1		1		
1	1						

3. KNIMEの画面からPMMLを見てみる—DMDを見てみる

datamodel.txt

```
use ic:施設型>ic:関与[都道府県名]>ic:関与者>ic:住所>ic:都道府県;  
use ic:施設型>ic:関与[市区町村名]>ic:関与者>ic:住所>ic:市区町村;  
use ic:施設型>ic:関与[市名町村コード]>ic:関与者>ic:ID>ic:識別値;  
データセット コードリスト  
use ic:施設型>ic:種別コード[災害種別]洪水>ic:識別値;  
use ic:施設型>ic:種別コード[災害種別]地震>ic:識別値;  
use ic:施設型>ic:種別コード[災害種別]崖崩れ、土石流及び地滑り>ic:識別値;  
use ic:施設型>ic:種別コード[災害種別]高潮>ic:識別値;  
use ic:施設型>ic:種別コード[災害種別]津波>ic:識別値;  
use ic:施設型>ic:種別コード[災害種別]大規模な火事>ic:識別値;  
use ic:施設型>ic:種別コード[災害種別]内水氾濫>ic:識別値;  
use ic:施設型>ic:種別コード[災害種別]火山現象>ic:識別値;  
use ic:施設型>ic:種別コード[指定緊急避難場所]>ic:識別値;  
use ic:施設型>ic:種別コード[指定避難所との重複]>ic:識別値;  
use ic:施設型>ic:種別コード[福祉避難施設所]>ic:識別値;  
use ic:施設型>ic:種別コード[その他の避難場所]>ic:識別値;  
use ic:施設型>ic:収容人数>ic:数値;  
use ic:施設型>ic:記述[搬入可能なトラックの大きさ]>ic:説明;
```

DMDの構造を見ると、表データの裏に隠れた

データセット

- 「災害種別」

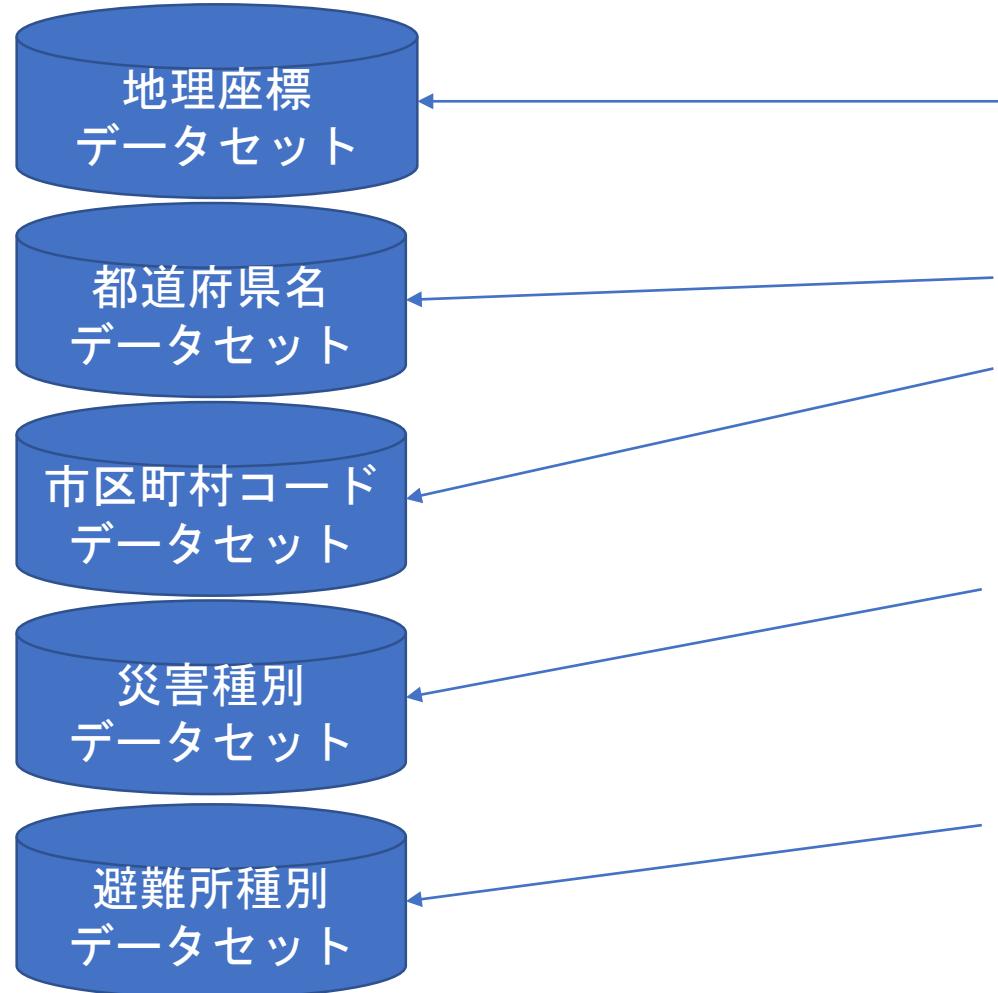
コードリスト

- 「洪水」
- 「地震」
- 「崖崩れ、土石流及び地滑り」
- 「高潮」
- 「津波」
- 「大規模な火事」
- 「内水氾濫」
- 「火山現象」

が見えてくる。

参照：00_変換元DMD(PD6943).zipのdatamodel.txt

表やDMD構造の裏には、データセットが隠れている



use ic:施設型>ic:ID>ic:識別値;
use ic:施設型>ic:名称>ic:表記;
use ic:施設型>ic:名称>ic:カナ表記;
use ic:施設型>ic:住所>ic:表記;
use ic:施設型>ic:住所>ic:方書;
use ic:施設型>ic:地理座標>ic:緯度;
use ic:施設型>ic:地理座標>ic:測地高度>ic:数値;
use ic:施設型>ic:地理座標>ic:経度;
use ic:施設型>ic:連絡先>ic:電話番号;
use ic:施設型>ic:連絡先>ic:内線番号;
use ic:施設型>ic:関与[都道府県名]>ic:関与者>ic:住所>ic:都道府県;
use ic:施設型>ic:関与[市区町村名]>ic:関与者>ic:住所>ic:市区町村;
use ic:施設型>ic:関与[市区町村コード]>ic:関与者>ic:ID>ic:識別値;
use ic:施設型>ic:種別コード[災害種別_洪水]>ic:識別値;
use ic:施設型>ic:種別コード[災害種別_地震]>ic:識別値;
use ic:施設型>ic:種別コード[災害種別_崖崩れ、土石流及び地滑り]>ic:識別値;
use ic:施設型>ic:種別コード[災害種別_高潮]>ic:識別値;
use ic:施設型>ic:種別コード[災害種別_津波]>ic:識別値;
use ic:施設型>ic:種別コード[災害種別_大規模な火事]>ic:識別値;
use ic:施設型>ic:種別コード[災害種別_内水氾濫]>ic:識別値;
use ic:施設型>ic:種別コード[災害種別_火山現象]>ic:識別値;
use ic:施設型>ic:種別コード[指定緊急避難場所]>ic:識別値;
use ic:施設型>ic:種別コード[指定避難所との重複]>ic:識別値;
use ic:施設型>ic:種別コード[福祉避難施設所]>ic:識別値;
use ic:施設型>ic:種別コード[その他の避難場所]>ic:識別値;
use ic:施設型>ic:収容人数>ic:数値;
use ic:施設型>ic:記述[搬入可能なトラックの大きさ]>ic:説明;
use ic:施設型>ic:記述[対象となる町会・自治会]>ic:説明;
use ic:施設型>ic:参照>ic:参照先;
use ic:施設型>ic:備考;
use ic:施設型>ic:ID[リンクID]>ic:識別値;

3. KNIMEの画面からPMMLを見てみる —予測モデルにDMDの構造を反映

```
-<DataField dataType="string" name="災害種別" optype="categorical">
  <Value value="洪水"/>
  <Value value="地震"/>
  <Value value="崖崩れ、土石流及び地滑り"/>
  <Value value="高潮"/>
  <Value value="津波"/>
  <Value value="大規模な火事"/>
  <Value value="内水氾濫"/>
  <Value value="火山現象"/>
</DataField>
```

データベースのテーブルの第3正規化とよく似ている！

3. KNIMEの画面からPMMLを見てみる —予測モデルにDMDの構造を反映

都道府県名を47都道府県のコードリストに
市区町村コードをコードリストに

```
-<DataField dataType="string" name="都道府県名" taxonomy="ic:施設型&amp;gt;
  <Value value="北海道"/>
  <!-- その他都道府県 -->
  <Value value="沖縄県"/>
</DataField>
+<DataField dataType="string" name="市区町村名" taxonomy="ic:施設型&amp;gt;
-<DataField dataType="string" name="市区町村コード" taxonomy="ic:施設型&amp;gt;
  <Value value="01100"/>
  <!-- その他都道府県コード -->
  <Value value="47382"/>
</DataField>
```

参照 : 02_pmml(項目を正規化).xml

避難所種別をコードリストに
参照 : 04_PD6943.xlsx (データ項目特記事項)

```
-<DataField dataType="string" name="指定緊急避難場所" taxonomy="ic:施設型&amp;gt;
  <Value value="1" displayValue="指定"/>
  <Value value="0" displayValue="指定していない"/>
</DataField>
-<DataField dataType="string" name="指定避難所" taxonomy="ic:施設型&amp;gt;
  <Value value="1" displayValue="指定"/>
  <Value value="0" displayValue="指定していない"/>
</DataField>
-<DataField dataType="string" name="福祉避難所" taxonomy="ic:施設型&amp;gt;
  <Value value="1" displayValue="指定"/>
  <Value value="2" displayValue="指定 (障害者等避難施設) "/>
  <Value value="3" displayValue="指定 (高齢者等避難施設) "/>
  <Value value="0" displayValue="指定していない"/>
</DataField>
-<DataField dataType="string" name="その他の避難場所" taxonomy="ic:施設型&amp;gt;
  <Value value="1" displayValue="指定"/>
  <Value value="0" displayValue="指定していない"/>
</DataField>
+<DataField dataType="double" name="想定収容人数" taxonomy="ic:施設型&amp;gt;
-<DataField dataType="string" name="搬入可能なトラックの大きさ" taxonomy="ic:施設型&amp;gt;
  <Value value="1" displayValue="4tまで可能"/>
  <Value value="2" displayValue="2tまで可能"/>
  <Value value="3" displayValue="乗用車サイズまで可能"/>
  <Value value="0" displayValue="指定していない"/>
</DataField>
```

3. KNIMEの画面からPMMLを見てみる —予測モデルにDMDの構造を反映

緯度経度の数値範囲を定める。

```
-<DataField dataType="double" name="緯度" taxonomy="ic:施設型&gt;ic:地理区  
  <Interval closure="closedClosed" leftMargin="-90.0" rightMargin="90.0"/>  
-<DataField dataType="double" name="標高" taxonomy="ic:施設型&gt;ic:地理区  
  <Interval closure="openOpen"/>  
-<DataField dataType="double" name="経度" taxonomy="ic:施設型&gt;ic:地理区  
  <Interval closure="closedClosed" leftMargin="-180.0" rightMargin="180.0"/>  
-<DataField>
```

参照 : 02_pmml(項目を正規化).xml

3. KNIMEの画面からPMMLを見てみる —予測モデルにDMDの構造を反映

Model - 0:1 - PMML Reader

File

No Table Spec – Columns: 12 Flow Variables

Columns: 12	Column Type	Column In...	Col...	Si...	Sh...	Filt...	Lower Bo...	Upper Bo...	Value 0	Value 1	Value 2	Value 3	Value 4	Value 5	Value 6	Value 7
緯度	Number (double)	0					-90	90	?	?	?	?	?	?	?	?
標高	Number (double)	1					0	0	?	?	?	?	?	?	?	?
経度	Number (double)	2					-180	180	?	?	?	?	?	?	?	?
都道府県名	String	3					?	?	北海道	沖縄県	?	?	?	?	?	?
市区町村コード	String	4					?	?	01100	47382	?	?	?	?	?	?
災害種別	String	5					?	?	洪水	地震	崖崩れ、土...	高潮	津波	大規模な火事	内水氾濫	火山現象
指定緊急避難場所	String	6					?	?	1	0	?	?	?	?	?	?
指定避難所	String	7					?	?	1	0	?	?	?	?	?	?
福祉避難所	String	8					?	?	1	2	3	0	?	?	?	?
その他の避難場所	String	9					?	?	1	0	?	?	?	?	?	?
想定収容人数	Number (double)	10					0	0	?	?	?	?	?	?	?	?
搬入可能なトラックの大きさ	String	11					?	?	1	2	3	0	?	?	?	?

機械が学習しやすい、予測データモデルになった

結論：予測モデルと語彙基盤

予測モデルと共通語彙基盤

位置付け

- 予測モデルを作りたい数学屋さん統計屋さんに、データのモデル（形・文脈）を伝える
- 入力データ設計、アルゴリズム・解析手法設計、学習手法の設計は、対象外！

期待すること

- データとDMDを公開しておくと、予測モデルに活用可能。標準的なデータマイニングツールに読み込む。
- DMDの構造を読んで、予測モデルを設計してもらう。
- データマイニングのユーザ層にアプローチしたい！

ありがとうございました！