



第2回コンテストについて

Index (第2回コンテストについて)

- 0. コンテストの概要について
- 1. 予測対象の高速道路について
- 1. 予測対象の高速道路について(続き)
- 2.1 「ドラぷら 高速料金・ルート検索サービス」について
- 2.2 ドラぷら検索ログデータについて
- 2.2.1 ドラぷら検索ログ2次データについて
- 2.2.1.1 ドラぷら検索ログ2次データ 構造化手順: 時間指定あり検索ログ
- 2.2.1.2 ドラぷら検索ログ2次データ 構造化手順: 時間指定なし検索ログ
- 3 提供データの仕様について
 - ▶ (1) 道路構造データ
 - ◆ 関越道及び東北道の道路構造データ
 - ◆ 全国道路構造データ
 - ▶ (2) トラカンデータ
 - ▶ (3) ドラぷら検索ログデータ
 - ◆ ドラぷら検索ログ1次データ
 - ◆ ドラぷら検索ログ2次データ
- 4. 道路構造データの仕様詳細
- 4.道路構造データの仕様詳細 (続き)
- 5. トラカンデータの仕様詳細
- 5.トラカンデータの仕様詳細 (続き)
- 6.1 ドラぷら検索ログ1次データの仕様詳細
- 6.1 ドラぷら検索ログ1次データの仕様詳細 (続き)
- 6.1.1 ルート検索補助プログラム
- 6.2 ドラぷら検索ログ2次データの仕様詳細
- 6.2 ドラぷら検索ログ2次データの仕様詳細 (続き)
- 7. 予測対象と予測タスクについて
- 7. 予測対象と予測タスクについて (続き)
- 8. 提出データの仕様
- 9. 評価基準について

コンテストの概要について

- 渋滞予測チャレンジコンテストにおいては、過去の交通量に関するデータ（交通量はトラカンと呼ばれるセンサーにより計測されています。）と高速料金・ルート検索のデータ（以下、ドラぷら検索ログ）から、**高速道路（関越道・東北道）のトラカン設置点における**（上り下りの双方）における渋滞の有無を予測します。
- 渋滞予測チャレンジコンテストにおいて、次の**3種類のデータ**を提供します。
 - ▶ 高速道路の構造（区間）について示したデータ
 - ▶ 高速道路の区間を挟むIC（インターチェンジ：高速道路の出入口）の名称、位置や当該区間に設置されたトラフィックカウンターの位置等の特徴量データなど
 - ▶ トラカンデータ
 - ◆ 高速道路に設置された計測機のデータで、走行車両の台数や速度等が含まれます。下記も参照してください。
<https://www.driveplaza.com/safetydrive/mamechishiki/029.html>
 - ▶ ドラぷら検索ログデータ
 - ◆ 高速道路のルート検索履歴に関するデータです。下記も参照してください。
<https://www.driveplaza.com/dp/SearchTop>

1. 予測対象の高速道路について

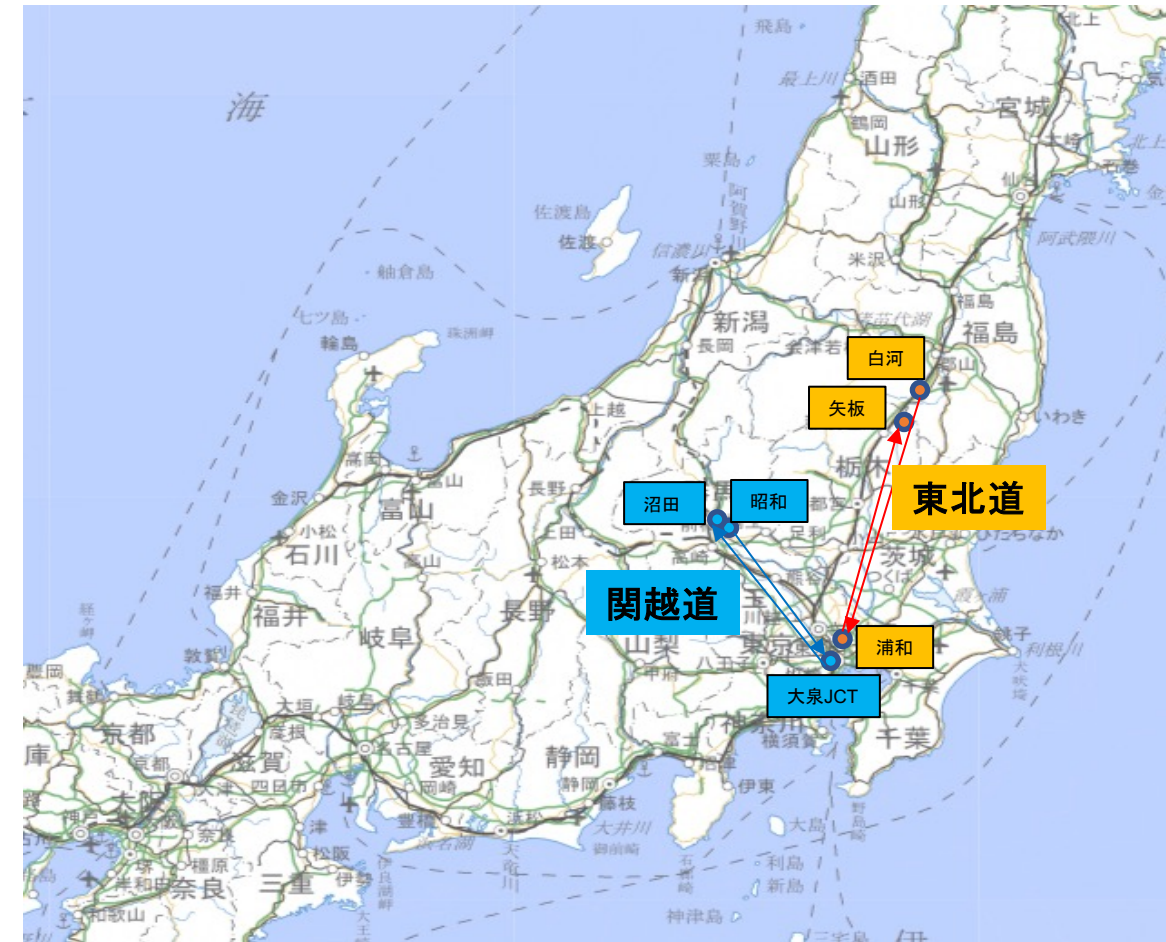
- 本コンテストでは、関越道及び東北道の渋滞を予測の対象とします。

■ 関越道

- 上り：「昭和」から「大泉JCT」
- 下り：「大泉JCT」から「沼田」

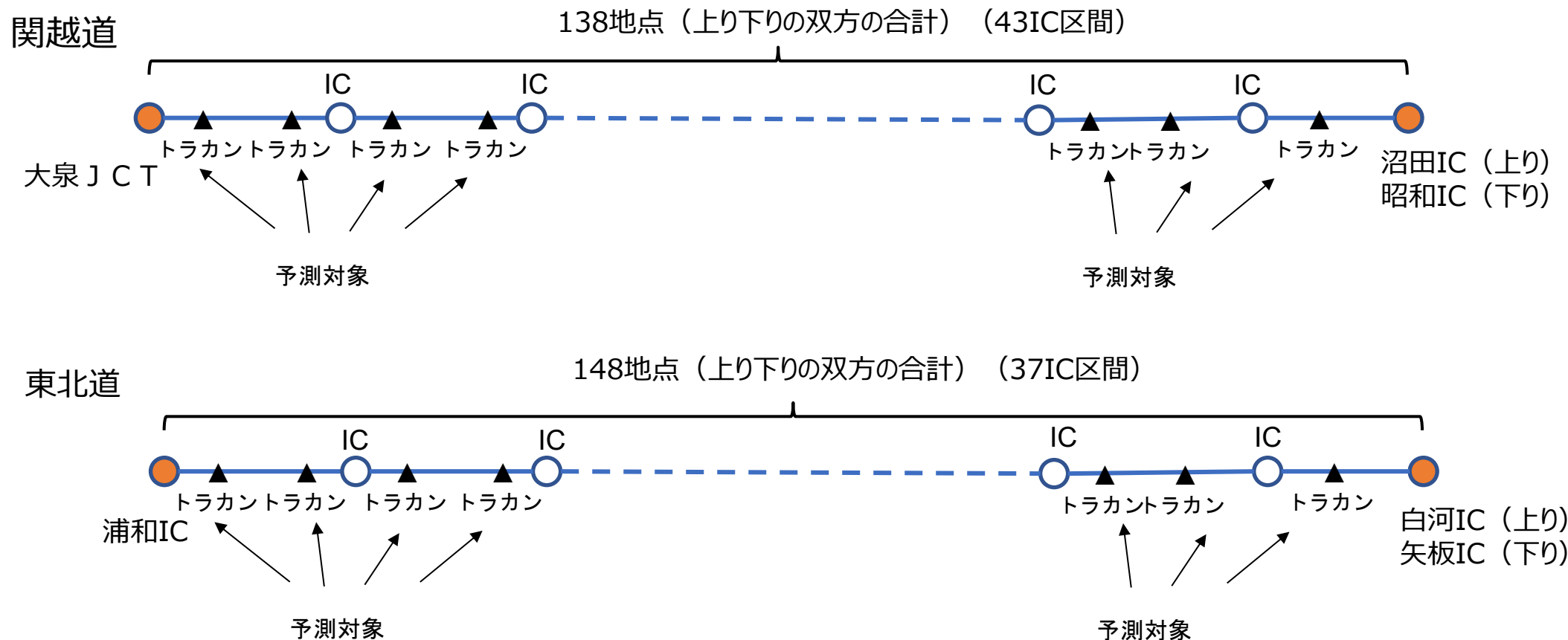
■ 東北道

- 上り：「白河」から「浦和」
- 下り：「浦和」から「矢板」



1. 予測対象の高速道路について(続き)

- 対象の高速道路の区間には、トラカンが数多く設置されており、本コンテストでは、その地点における渋滞の有無の予測を行います。第1回のコンテストでは、IC区間を予測対象としていましたが、今回は、より細かいトラカンの設置点における渋滞の予測を行うことになっていますので、留意して下さい。



2.1 「ドラぷら 高速料金・ルート検索サービス」について

- 検索の流れ
 - ▶ ルートの出発・到着インターチェンジ（IC）を指定する
 - ◆ ルート中の経由ICを指定可能（本渋滞予測チャレンジコンテンツでは考慮しない）
 - ◆ 検索するルートを使用する予定日時を指定可能
 - ◆ 指定した日時が出発時なのか到着時なのかを指定可能
- ドラぷら 料金・ルート検索1件ごとにログ（以降「ドラぷら検索ログ」）が1行記録される（以下5つの項目）
 - ▶ 検索日時
 - ◆ 料金・ルート検索が実行された日時
 - ▶ 出発IC
 - ◆ 高速道路に入るIC
 - ▶ 到着IC
 - ◆ 高速道路から出るIC
 - ▶ 指定日時
 - ◆ 検索されたルートを移動する予定の日時
 - ◆ 指定されない場合はnullがセットされる
 - ▶ 指定種別
 - ◆ 指定日時が出発時か到着時か
 - ◆ 指定されない場合はnullがセットされる

ドラぷら 高速料金・ルート検索サービス トップページ

「地図からIC名を調べて入力」または「住所やキーワードなどから入力」をクリックすると「地図」「住所・キーワード」「道路名」「履歴」いずれかの方法でIC名を調べられます。

出発IC: 練馬 (地図からIC名を調べて入力 / 住所やキーワードなどから入力)

到着IC: 湯沢(関越道) (地図からIC名を調べて入力 / 住所やキーワードなどから入力)

日時: 2023/03/01 10:00 (出発時 / 到着時)

検索する

ルートの出発・到着ICを指定

ルートを使う予定の日時を指定
(指定しない場合は null が入る)

検索日時	出発IC	到着IC	指定日時	指定種別
2023/mm/dd HH:MM	練馬	湯沢	2023/3/1 10:00	出発

2.2 ドラぷら検索ログデータについて

- 本コンテストでは、「ドラぷら 高速料金・ルート検索サービス」から得られたデータとして、以下2種類のデータを提供します。
 - ドラぷら検索ログ2次データ
 - 参加者の利便性のために、検索データを、時刻、高速道路の区間及びトラカンと関連した形で提供します。（第1回と同様）
 - ドラぷら検索ログ1次データ
 - より高度な予測を可能とするために、ドラぷら検索ログのデータをそのままに近い形で提供します。便宜上、「ドラぷら検索ログ1次データ」と呼んでいますが、コンテストのために若干の加工を行って提供しています。また、その検索ログに含まれる出発地点及び到着地点から経路を探索プログラムを提供します。

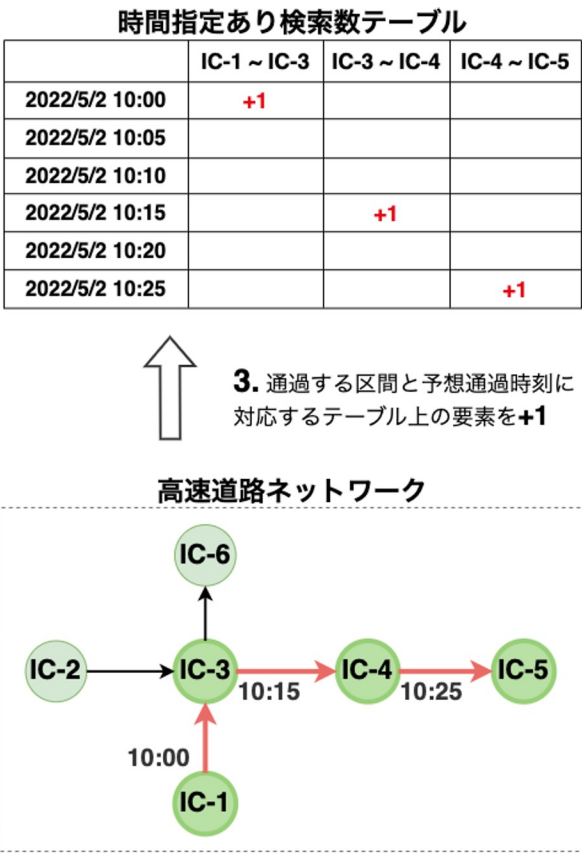
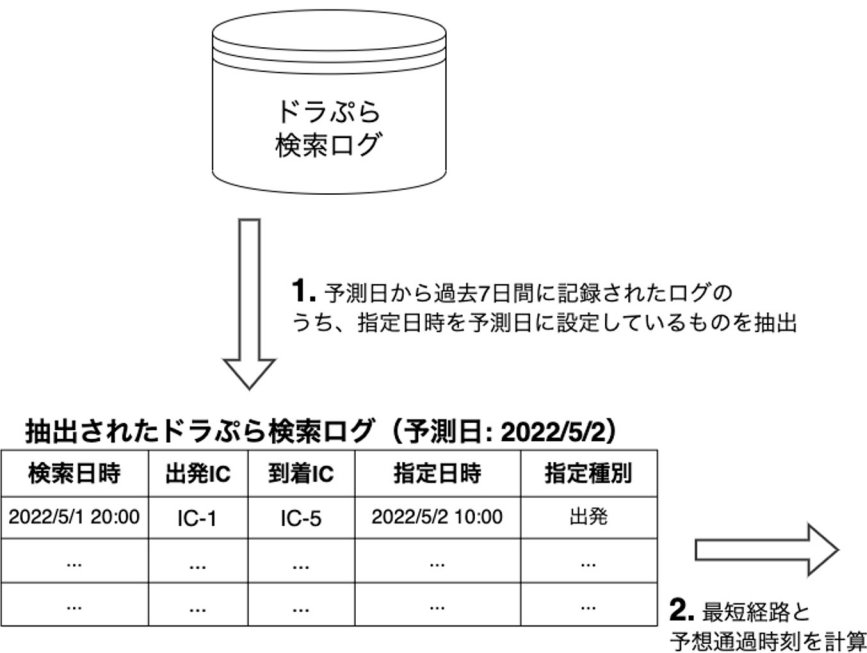
2.2.1 ドラぷら検索ログ2次データについて

- 本渋滞予測チャレンジコンテストでは、コンテストの利便性のために、トラカンデータと関連する形でドラぷら検索ログデータの構造化を行い、構造化後のドラぷら検索ログを**ドラぷら検索ログ2次データ**として提供する
- ドラぷら検索ログは渋滞予測の入力データとして活用しづらいため構造化する
 - ▶ ドラぷら検索ログからは「検索されたルートがどの区間を通過するのか」という詳細な経路情報や「その区間をいつ通過するのか」という時間情報を直接取り出すことができない
- 構造化の際、ドラぷら検索ログを2種類に分割
 - 1. 時間指定あり検索ログ**: ドラぷらの料金・ルート検索時に移動日時を指定しているログ
 - ◆ より直接的に将来の潜在的な交通量として解釈できる
 - 2. 時間指定なし検索ログ**: ドラぷらの料金・ルート検索時に移動日時を指定していないログ (= 指定日時がnull)
 - ◆ 将来のどのタイミングで交通状況に影響を及ぼすかは分からないが、将来の交通状況と何らかの相関が存在すると仮定
- 時間指定あり検索ログ、時間指定なし検索ログを別の方法で構造化し、その結果の値を「**時間指定あり検索数**」「**時間指定なし検索数**」としてそれぞれ定義する (2.2.1.1および2.2.1.2を参照)

2.2.1.1 ドラぷら検索ログ2次データ 構造化手順: 時間指定あり検索ログ

■ 時間指定あり検索ログを以下の手順で時間指定あり検索数テーブルの形に構造化する

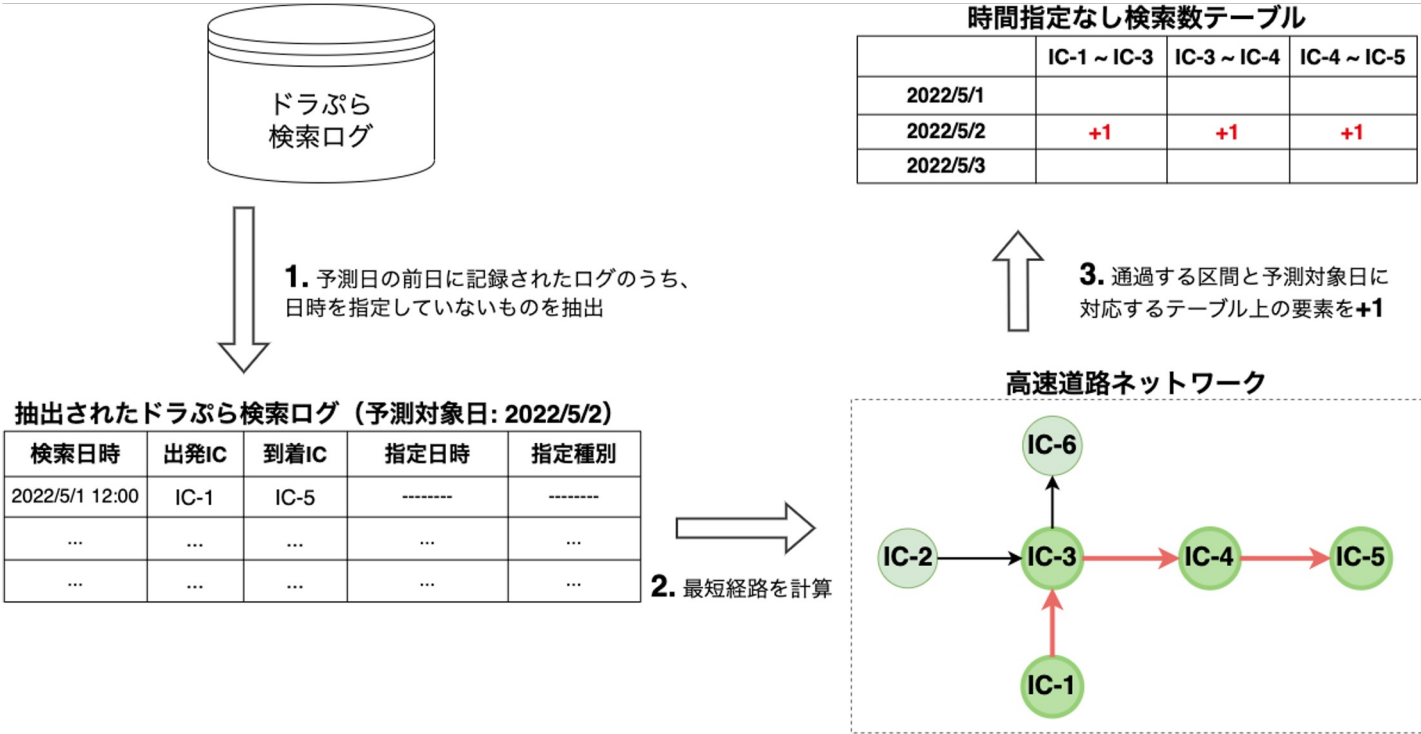
- 1. 予測対象日から**過去7日間**に記録されたドラぷら検索ログのうち、**指定日時が予測対象日に設定されているログのみ**を抽出
- 2. 抽出されたログの**出発ICから到着ICに至る最短経路**と、経路中の**各区間の予想通過時刻**を計算
 - ▶ 各区間の予想通過時刻は**5分単位**に丸められる
- 3. 時間 × ICの2次元で構成される「**時間指定あり検索数テーブル**」を用意し、「**5分単位に丸められた時刻に各区間を通過するルートを検索したログの数**」をテーブル上の対応する要素に加算していく



2.2.1.2 ドラぷら検索ログ2次データ 構造化手順: 時間指定なし検索ログ

■ 時間指定なし検索ログを以下の手順で時間指定なし検索数テーブルの形に構造化する

1. 予測対象日の**前日**に記録されたドラぷら検索ログのうち、**日時を指定していないログのみ**を抽出
2. 抽出されたログの**出発ICから到着ICに至る最短経路**を計算
3. 時間 × ICの2次元で構成される「**時間指定なし検索数テーブル**」を用意し、「**各区間を通過するルートを検索したログの数**」を**対応する区間ごとに予約対象日 (= 検索実行日の翌日) の行に加算**していく
 - ▶ 日時を指定せずに検索されたルートは将来的な移動のタイミングが不明であるため、時間指定あり検索よりも粒度を粗くする
 - ▶ そのため、各区間について、「その区間を通過するルートを検索したログ数」を**1日ごとに集計**している



3. 提供データの仕様について

(1) 道路構造データ

- ◆ 関越道及び東北道の道路構造データ (road_local.csv)

- ◆ 関越道・東北道における各区間の始点、終点や長さ、距離などの属性データです。渋滞予測を行う際の基本となるデータを提供します。

- ▶ 主キー

- ◆ 道路区間の始点コード `start_code`
- ◆ 道路区間の終点コード `end_code`

- 全国道路構造データ (road_all.csv)

- 関越道・東北道を含む全国道路における各区間の始点、終点や方向、距離などの属性データです。ドラぷら検索ログと合わせて使用します。

- ▶ 主キー

- ◆ 道路区間の始点コード `start_code`
- ◆ 道路区間の終点コード `end_code`

3. 提供データの仕様について（続き）

(2) トラカンデータ（train.csvとtest.csvに分割されています。）

- ▶ 各道路区間を通過する自動車の速度、台数などを記録した時系列データ
- ▶ 主キー
 - ◆ 日時 `datetime`
 - ◆ 道路区間の始点コード `start_code`
 - ◆ 道路区間の終点コード `end_code`
 - ◆ トラフィックカウンターが位置するキロポストの値 `KP`
- ▶ データ期間
 - ◆ コンテスト期間に使用可能なデータ
 - ・ 学習用（train.csv）：2021年4月8日-2023年7月31日（845日分の教師データを配布します。）
 - ・ 暫定評価用（test.csv）：2023年8月1日-2023年9月30日（参加者には配布されません。暫定評価に使用します。）
 - ◆ 最終評価期間に使用可能なデータ
 - ・ 推論補助用：2023年10月1日-2024年3月30日（参加者には配布されません。最終評価で再学習に使用可能。）
 - ・ 最終評価用：2024年3月31日-2024年5月6日（参加者には配布されません。最終評価に使用します。）

3. 提供データの仕様について（続きの2）

(3) ドラぷら検索ログデータ

■ ドラぷら検索ログ 1 次データ (search_raw_log/yyyy_mm_dd.csv)

- 「ドラぷら」から収集された検索履歴のうち、検索された経路が関越道または東北道を通過するもの

◆ 主キー

- 日時 `datetime`
- 道路区間の始点コード `start_code`
- 道路区間の終点コード `end_code`

■ ドラぷら検索ログ 2 次データ

▶ 各道路区間の時間指定あり検索数 (search_specified.csv)

- ▶ 構造化の結果得られた時間指定あり検索数テーブルの値を（時間方向に）1時間単位でリサンプリングした値

◆ 主キー

- 日時 `datetime`
- 道路区間の始点コード `start_code`
- 道路区間の終点コード `end_code`

▶ 各道路区間の時間指定なし検索数 (search_unspecified.csv)

- ▶ 前述の構造化の結果得られた時間指定なし検索数テーブルの値

◆ 主キー

- 日付 `date`
- 道路区間の始点コード `start_code`
- 道路区間の終点コード `end_code`

4. 道路構造データの仕様詳細

道路構造データは、高速道路の区間を挟むICの名称、位置や当該区間に設置されたトラフィックカウンターの位置等を示す、静的なデータであり詳細は次のとおり。

■ 1. 全国のICネットワークデータ(road_all.csv)

データ仕様				
ヘッダ名称	データ型	説明	値例	備考
start_name	str	区間始点IC名称	所沢	
end_name	str	区間終点IC名称	大泉 J C T	
start_code	str	区間始点ICコード	1800006	
end_code	str	区間終点ICコード	1110210	
direction	str	方向	上り;下り	上り (0kpへ向かう方向) 下り (0kpから離れる方向)
distance	float64	区間距離	5.0	5
road_code	str	道路コード	1800	

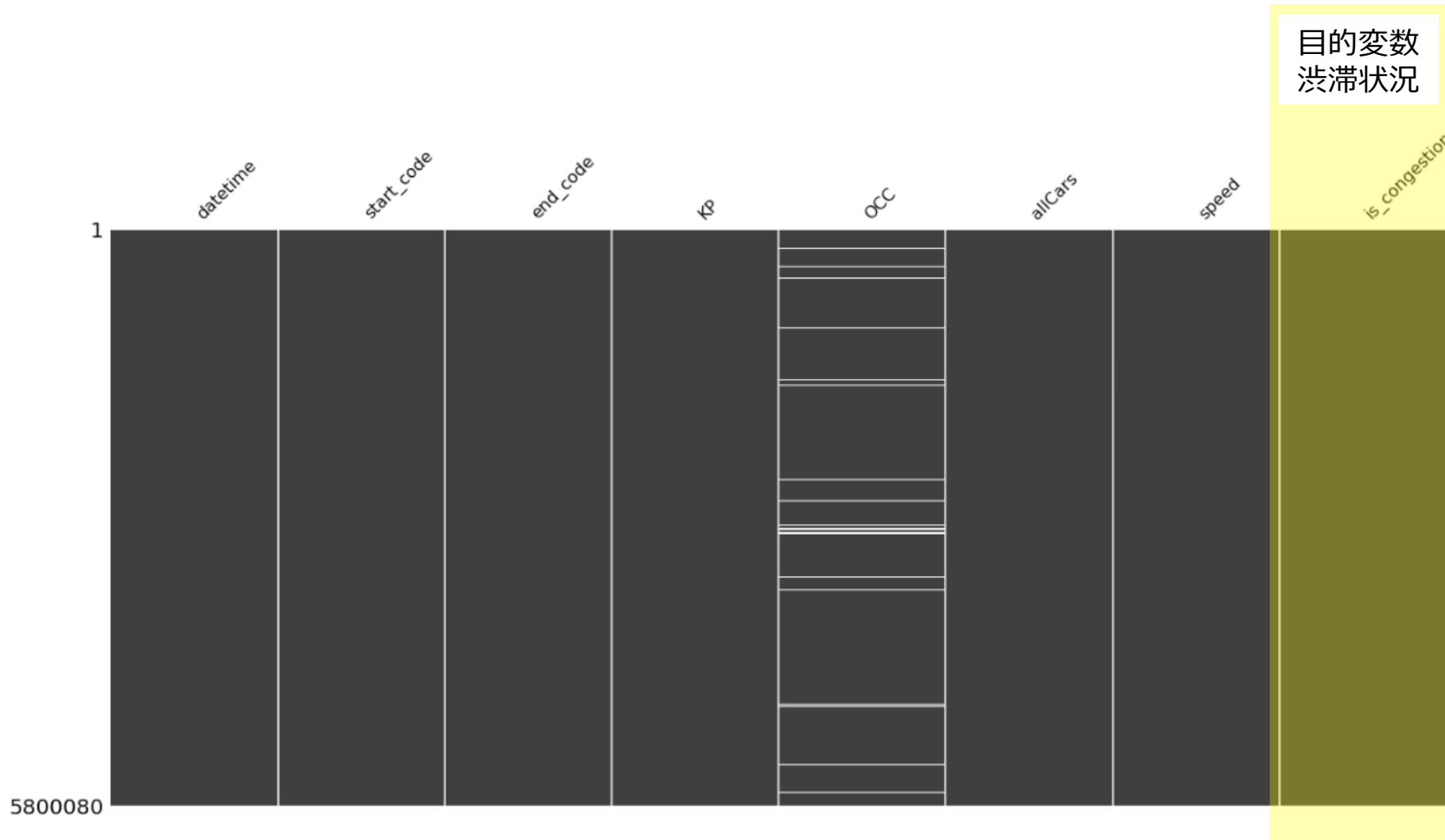
4. 道路構造データの仕様詳細（続き）

▶ 2. 関越・東北に限定した道路構造データ及び区間における特徴量データ(road_local.csv)

データ仕様				
ヘッダ名称	データ型	説明	値例	備考
start_name	str	区間始点IC名称	所沢	
end_name	str	区間終点IC名称	大泉 J C T	
start_code	str	区間始点ICコード	1800006	
end_code	str	区間終点ICコード	1110210	
start_pref_code	int64	区間始点ICが位置する県コード	11	https://nlftp.mlit.go.jp/ksj/gml/codelist/PrefCd.html
end_pref_code	int64	区間終点ICが位置する県コード	13	
start_lat	float64	区間始点ICの緯度	35.80615	
end_lat	float64	区間終点ICの緯度	35.75582	
start_lng	float64	区間始点ICの経度	139.535511	
end_lng	float64	区間終点ICの経度	139.601514	
start_degree	float64	区間始点ICの度数	2	各ICが接続する他ICの数
end_degree	float64	区間終点ICの度数	4	
direction	str	方向	上り;下り	上り（0kpへ向かう方向） / 下り（0kpから離れる方）
start_KP	float64	始点のICを代表するキロポスト（km）	43.7	<ul style="list-style-type: none"> KP（キロポスト）は、起点を0kpとして、道路中心線に沿う位置。 関越道起点:「大泉 J C T（練馬IC）」 東北道は「川口JCT」を起点としているため、「浦和」のKPは35.7となる。
end_KP	float64	終点のICを代表するキロポスト（km）	35.7	
limit_speed	float64	制限速度（km/h）	100	
road_code	int64	道路コード	1800	1800（関越道）;1040（東北道）

5. トラカンデータの仕様詳細

- データ粒度：1時間間隔
- 学習用データ (train.csv)の合計：5800080（行）× 8（列）



5.トラカンデータの仕様詳細（続き）

■ 学習用データ (train.csv)

	ヘッダ名称	データ型	説明	値例	備考
時刻	datetime	str	時刻	2021-04-08 00:00:00	
トラフィックカウンター情報	start_code	str	区間始点ICコード	1800006	道路構造データを参照
	end_code	str	区間終点ICコード	1110210	
	KP	float64	トラフィックカウンターが位置するキロポストの値		
トラカンデータ	allCars	int64	全車線の合計通過台数	568	計測器から速度、台数、OCCが計測されている
	OCC	float64	全車線の占有率（%）	1.6	
	speed	float64	全車線の平均速度	87	
目的変数	is_congestion	{0,1}	渋滞有無フラグ	1/0	1：渋滞あり/ 0:渋滞なし

6.1 ドラぷら検索ログ1次データの仕様詳細 (search_raw_log/yyyy_mm_dd.csv)

- 以下のディレクトリ構造のように日付ごとにデータを分割して配布

search_raw_log
├── 2021_04_01.csv
├── 2021_04_02.csv
├── ...
├── 2023_07_30.csv
└── 2023_07_31.csv

データ仕様				
ヘッダ名称	データ型	説明	値例	備考
datetime	str	検索日時	2021-04-01 00:00:10	
start_code	str	区間始点ICコード	1040263	
end_code	str	区間終点ICコード	1010046	
spec_datetime	float	指定日時	2021-04-01 07:00:00	上り (0kpへ向かう方向) 下り (0kpから離れる方向)
spec_type	int64	指定日時の種別 (出発時 or 到着時)	1 or 2	1: 出発時 ; 2: 到着時
car_type	int64	車両種別	{1,2,3,4,5}	1: 軽自動車等 ; 2: 普通車 ; 3: 中型車 ; 4: 大型車 ; 5: 特大車

6.1.1 ルート検索補助プログラム (expressway_router.py) について

- ドラぷら検索ログ1次データを活用しやすいように、出発ICから到着ICまでに通過する経路を計算するためのプログラムを提供します。
- 本プログラムでは、出発・到着ICのICコードを入力すると、経路中に通過するIC一覧がICコードの配列として返されます。
- 使い方の詳細については、 `test_expressway_router.ipynb` を参照してください。

6.2 ドラぷら検索ログ2次データの仕様詳細

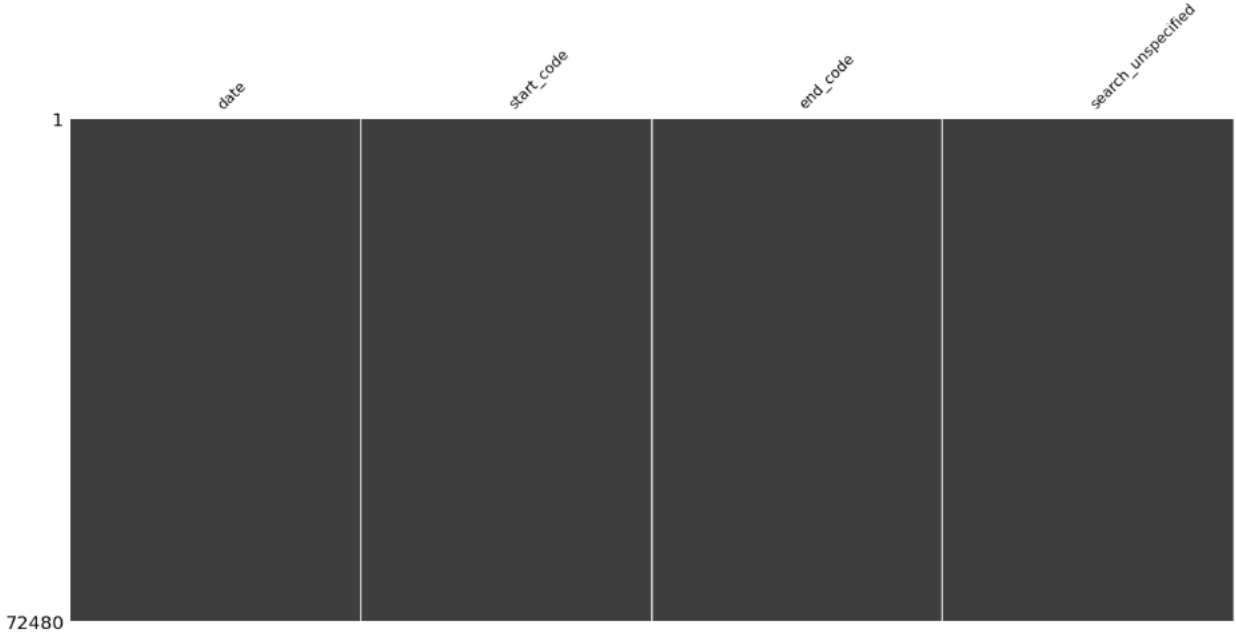
- 時間指定あり検索数（search_specified.csv）
- データサイズ
 - ▶ 1739520（行） × 4（列）

1	datetime	start_code	end_code	search_specified
1739520				

ヘッダ名称	データ型	説明	値例	備考
datetime	str	時刻	2021-04-08 00:00:00	
start_code	str	区間始点ICコード	1800006	道路構造データを参照
end_code	str	区間終点ICコード	1110210	
search_specified	float64	ドラぷら検索ログ（時間指定あり 検索数（1時間単位）	15.0	詳細は「ドラぷら検索ログデータ」を参照

6.2 ドラぷら検索ログ2次データの仕様詳細（続き）

- 時間指定なし検索数（search_unspecified.csv）
- データサイズ
 - ▶ 72480（行） × 4（列）



ヘッダ名称	データ型	説明	値例	備考
date	str	日付	2021-04-08	
start_code	str	区間の始点コード	1800006	
end_code	str	区間の終点コード	1110210	
search_unspecified	float64	ドラぷら検索ログ（時間指定なし検索数（過去1日単位の検索数）	2142.0	詳細は「ドラぷら検索ログデータ」を参照

7. 予測対象と予測タスクについて

■ 予測対象（渋滞の有無）について

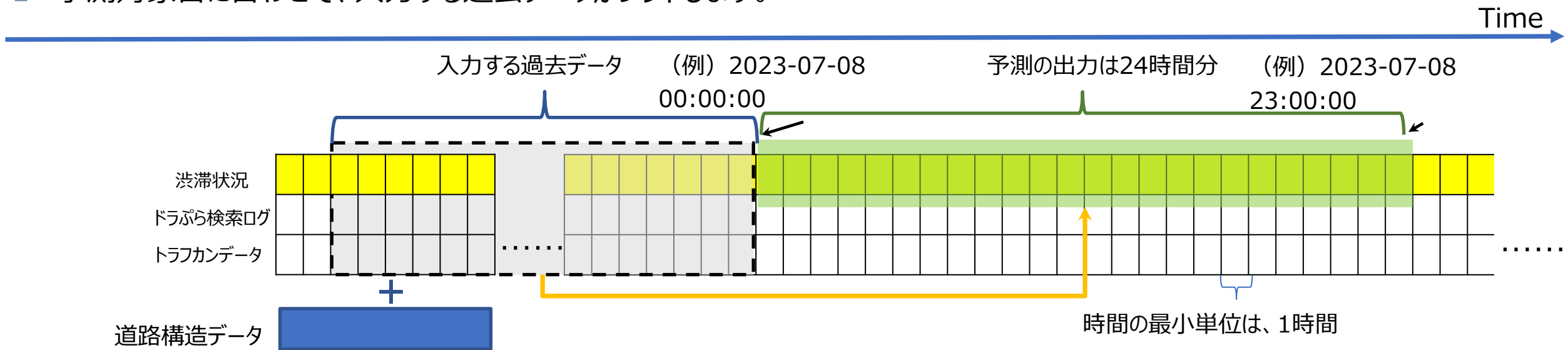
- ▶ 予測の対象とする渋滞の有無については、各区分（上り下りの両方）に設置されたトラフィックカウンターから計測された自動車の平均速度が時速40km以下の場合に、渋滞が発生している（渋滞有り）と定義します。
- ▶ なお、上記渋滞の定義は、NEXCO東日本における渋滞の定義とほぼ一致しますが、厳密な渋滞の定義とは異なり、渋滞予測チャレンジコンテストのため、便宜的に用意したものです。




■ 予測タスクについて

- ▶ 過去のトラカンデータ及びドラぷら検索ログ並びに道路構造データ（input）から、翌日の0時から24時までの高速道路（関越道・東北道）の各トラフィックカウンター設置地点における渋滞の有無（output）を、1時間毎に予測します。
- ▶ 予測の対象範囲は、次のとおりとします。
 - ◆ 暫定評価期間：2023年8月1日～2023年9月30日の週末、祭日、連休等（下記の指定日のとおり。）
 - 2023年8月5日、2023年8月6日、2023年8月11日～2023年8月20日、2023年8月26日、2023年8月27日、2023年9月2日、2023年9月3日、2023年9月9日、2023年9月10日、2023年9月16日、2023年9月17日、2023年9月18日
 - ◆ 最終評価期間：2024年4月1日～2024年5月6日の指定日（下記の指定日のとおり。）
 - 2024年4月6日、2024年4月7日、2024年4月13日、2024年4月14日、2024年4月20日、2024年4月21日、2024年4月26日～2024年5月6日

7. 予測対象と予測タスクについて（続き）

- 予測タスクにおける入力と出力の関係は次の通り。
- 予測対象日に合わせて、入力する過去データがシフトします。



	説明
	入力する固定データ ・道路構造データは、固定値です。
	入力する過去データ ・過去データ（トラカンデータ及びドラぷらルート検索データ）の長さは任意です。
	出力データ ・毎日0時に予測タスクを行うことを想定しています。 ・0時において、次の24時間分（24個）の渋滞状況を出力し、予測対象日数に合わせて、予測を繰り返します。

8. 提出データの仕様について

- 提出データの仕様について

予測プログラムが提出物となります。

提出されたプログラムを用いて、SIGNATEサーバーにおいて予測を行い、そして評価を行います。

▶ 詳細は配布データのreadme.mdを参照してください。

9. 評価基準について

- 渋滞予測チャレンジコンテストの評価は、評価対象期間の全区間（上り下りの両方）における予測結果（渋滞の有無）を対象に、次の指標により行います。
- 評価指標について（精度賞）
 - ▶ 評価期間中の全ての渋滞（正解）と予測のF1スコアを評価指標とします。

$$F1Score = \frac{2}{\frac{1}{precision} + \frac{1}{recall}}$$

	予測：渋滞あり	予測：渋滞なし	
正解：渋滞あり	True Positive (TP) 正解は渋滞あり 予測は渋滞あり	False Negative (FN) 正解は渋滞あり 予測は渋滞なし	$Recall = \frac{TP}{TP + FN}$
正解：渋滞なし	False Positive (FP) 正解は渋滞なし 予測は渋滞あり	True Negative (TN) 正解は渋滞なし 予測は渋滞なし	

$Precision = \frac{TP}{TP + FP}$

- 評価指標について（モデリング賞）
 - ▶ 精度賞受賞者の内、提出したモデル内容・表彰式における発表により、審査員が定性的に評価を行います。