기술의 원리에 대해 고민하는 개발자입니다.

김선중 010-2405-7290 govidnah08@gmail.com

CONTENTS

- 1. <u>자기소개</u>
- 2. <u>기술 보유</u>
- 3. <u>학력 및 경력</u>
- 4. <u>참여 프로젝트</u>
- 5. 프로젝트
 - 5.1 신경망의 Lipschitz constant 계산 (석사논문)
 - 5.2 <u>신규 쏘카존 제안 (AIFFEL)</u>
 - 5.3 신호생성 모듈 및 통행배정 모듈 개발 (성남시 ITS 사업)

1. 자기소개



Sun Joong Kim

수학적인 사고능력

제 전공(학부, 대학원)은 수학입니다. 수학을 오래 공부한 만큼 주어진 문제를 수학적으로 해석하고 활용하는 데에 특기가 있습니다.

최적화 문제에 대한 해결

수학 전공을 가장 잘 발휘할 수 있는 응용분야 중 하나는 최적화(optimization)입니다. 다양한 문제상황을 최적화 문제로서 적절히 정의하고 선형계획법, 유전알고리즘, 신경망 등을 활용하여 최적화문제를 해결할 수 있습니다.

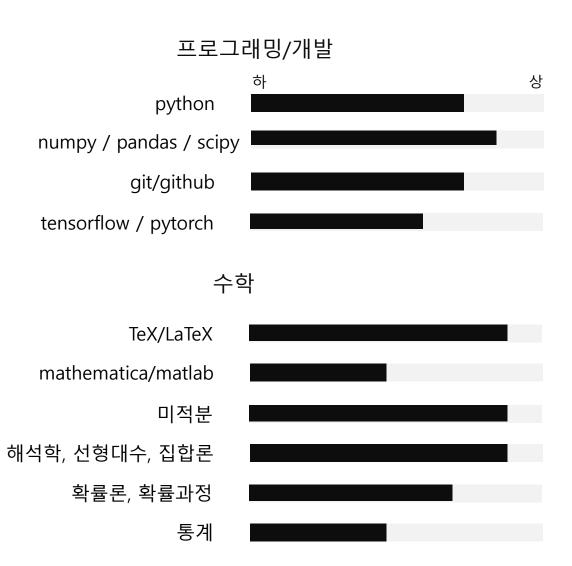
머신러닝/딥러닝 이론에 대한 이해

대학원 재학 중 지도교수님의 지도 아래에서 머신러닝/딥러닝에 관한 이론을 폭넓게 공부했습니다. MLP, CNN, RNN, attention, transformer, auto-encoder, GAN 등에 대한 지식을 가지고 있습니다. 이외에, 강화학습, GNN(graph NN), BNN(Bayesian NN), NLP(natural language processing)에 대해서도 간략하게 익혔습니다.

2. 기술 보유

기술 보유

- python 등 프로그래밍 개발 가능
- numpy/pandas를 통한 데이터 처리 가능
- git/github를 통한 협업 가능
- keras/tensorflow를 통한 인공지능 아키텍쳐 구축 가능
- TeX/LaTeX을 통한 수식 문서 작성 능숙



3. 학력 및 경력



Sun Joong Kim

EDUCATION

2007. 02 춘천고등학교 졸업 2008. 03 고려대학교 수학과 학부 입학 2015. 03 고려대학교 수학과 학부 졸업 2019. 08 고려대학교 수학과 대학원 입학 2023. 02 AIFFEL 쏘카 3기 수료 2024. 02 고려대학교 수학과 대학원 졸업

LICENSE

데이터분석 준전문가(ADsP) | 2022.11 취득

CAREER

2014 ~ 2018 수학 개인 교습(전업) 2021 ~ 2022 IBEDU 온라인 수학 개인 교습 강사 2023 ~ 2024 유아이네트웍스 교통연구본부 대리

PROJECT

2020 LSTM/wavelet을 활용한 주식 예측 (대학원)2022 신경망의 Lipschitz constant 계산 (대학원 석사논문)2023 SOCAR, 신규 쏘카존 제안 프로젝트 (AIFFEL)2024 신호생성 및 통행배정 모듈 개발 (성남시 ITS 사업)

4. 참여 프로젝트

프로젝트명	프로젝트 기간	설명				
LSTM/wavelet을 활용한 주식예측 연구(대학원)	2020.07 ~ 2020.12	- wavelet을 활용한 주식 추세선 분해 - LSTM을 통한 주식 예측				
신경망의 Lipschitz constant 계산 (석사논문)	2022.07 ~ 2022.09	- 인공신경망의 Lipschitz constant 계산법 정리한 survey paper - 해당 상수는 알고리즘의 안정성(robustness와 연관) - 일변수함수, 선형함수에 대한 Lipschitz constant 계산법 - 신경망(neural network), CNN에 대한 Lipschitz constant 계산법				
신규 쏘카존 제안 프로젝트(AIFFEL)	2022.12 ~ 2023.02	- 경기도 지역에 대한 카셰어링 기반 군집화 - 군별 수요예측(SARIMA) - 정량적 방법을 통한 쏘카존 제안				
성남시 ITS 사업 (UI networks)	2023.09 ~ 2024.09(예정)	- 실시간 신호생성 모듈 개발 - 통행배정 모듈 개발				

5. 프로젝트

5.1 신경망의 Lipschitz constant 계산 (석사논문)

프로젝트명	신경망의 Lipschitz constant 계산
프로젝트기간	2022.07~2022.09
프로젝트인원	개인
설명	머신러닝/딥러닝 모델의 안정성(robustness)와 관련되어 있는 립시 츠 상수(Lipschitz constant)의 계산법에 관한 연구 논문 작성.
담당역할	선행연구 분석 및 관련개념 정리, 논문 작성
링크	https://github.com/govin08/Lipschitz Constant of DNN/blob/main/Lipschitz constant of neural networks.pdf
비고	대학원 석사 논문

Master's Thesis

Lipschitz Constants of Functions of Neural Networks

Sunjoong Kim

Department of Mathematics

Graduate School

Korea University

February 2024

[그림 6-1. 논문 표지]

Chapter 1. Introduction

어떤 머신러닝/딥러닝 모델이 좋은 성능을 내느냐 하는 문제만큼이나 중요한 것은 해당 모델이 안정적(robust)으로 동작하느냐 하는 문제입니다. 모델의 안정성과 관련있는 수학적인 값은 립시츠 상수(Lipschitz constant)입니다. 유클리드 공간 사이에서 정의된 함수에 대한 립시츠 상수의 정의는 오른쪽 그림과 같습니다.

예를 들어, GAN의 한 종류인 Wasserstein GAN은 1-Lipschitz function만을 고려함으로써, GAN의 고질적인 문제인 불안정성을 해결한 바 있습니다.

Definition 1. A function $f: \mathbb{R}^n \to \mathbb{R}^m$ is called Lipschitz continuous if there is a nonnegative real number c such that

$$||f(x) - f(y)|| \le c||x - y|| \qquad (x, y \in \mathbb{R}^n)$$
 (2.1.2)

The infimum of c which satisfies (2.1.2) is called the Lipschitz constant of f and is denoted by Lip(f).

[립시츠 상수의 정의]

Chapter 2. Lipschitz constant of functions between Euclidean spaces

머신러닝/딥러닝에서 자주 사용되는 기본 함수들(활성화함수, 선형함수 등)에 대한 립시츠 상수를 계산해봅니다. 또한 함수들의 합성이 립시츠 상수에 어떤 영향을 미치는지 알아봅니다.

선형함수에 대한 립시츠 상수는 특히 중요합니다. 그것은 그 선형함수에 대응되는 행렬에 대한 작용소 놈(operator norm)과 일치합니다. 행렬의 작용소 놈에 대한 계산방법이 설명됩니다.

Table 2.1: The Lipschitz constants of univariate activation functions.

Activation Functions	Formula	Lip(f)
Sigmoid	$\sigma(x) = \frac{1}{1 + e^{-x}}$	1/4
Hyperbolic tangent	$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	1
Rectified Linear Unit	$ReLU(x) = max\{0, x\}$	1
Leaky ReLU	$LReLU(x) = max\{\alpha x, x\}$	1
Exponential Linear Unit	$ELU(x) = \begin{cases} x & (x \ge 0) \\ \alpha(e^x - 1) & (x < 0) \end{cases}$	1
Softplus	$f(x) = \frac{1}{\beta} \log(1 + e^{\beta x})$ $g(x) = e^{-x^2}$	1
Gaussian	$g(x) = e^{-x^2}$	$\sqrt{\frac{2}{e}}$

[활성화함수들의 립시츠 상수]

2.3 Linear Functions

2.3.1 Operator Norms ||W||

Consider a linear function f(x) = Wx for some $m \times n$ matrix W. Because of the linearity, the condition (2.1.2) reduces to

$$||Wx|| \le c||x||.$$
 $(x \in \mathbb{R}^n)$ (2.3.14)

The smallest c which satisfies (2.3.14) is called *the operator norm of W* and denoted by ||W||. Thus, the optimal Lipschitz constant of f equals the operator norm of W; ||W|| = Lip(f). The definitions of the forms (2.1.4) are as follows;

[선형함수의 립시츠 상수]

Chapter 3. Lipschitz constants of neural networks 신경망에서의 립시츠 상수를 계산해봅니다. 구체적으로는 MLP와 CNN에 대한 립시츠 상수를 계산해봅니다.

하지만 2-layered MLP만 되더라도 정확한 립시츠 상수를 구하는 것이 불가능 (NP-hard)하기 때문에 립시츠 상수의 한 upper bound를 구하는 알고리즘을 소개합니다(AutoLip).

CNN에 대해 계산할 때에는 가로로 긴 행렬의 행렬곱을 생각하게 됩니다. 이 때 계산을 편하게 하기 위한 방법으로 Power method와 Rayleigh quotient 등을 사용할 수 있습니다.

```
input: the function f = g_K \circ \cdots \circ g_0, where \theta_k = (g_k \circ \cdots \circ g_0)(x) for 0 \le k \le K.

output: An upper bound L for \mathrm{Lip}(f)

1 k \leftarrow 0;

2 L_0 \leftarrow 1;

3 while k < K do

4 \left| \begin{array}{c} L_k \leftarrow \sum_{i=0}^{k-1} \left( \sup \left| \left| \frac{\partial \theta_k}{\partial \theta_i} \right| \right| \right) L_i;

5 \left| \begin{array}{c} k \leftarrow k + 1 \end{array} \right|;

6 end

7 L \leftarrow L_K
```

Algorithm 1: AutoLip

[그림 6-5. AutoLip]

```
input: A rectangular matrix F \in \mathbb{R}^{n \times m}.
output: Approximation for ||F||

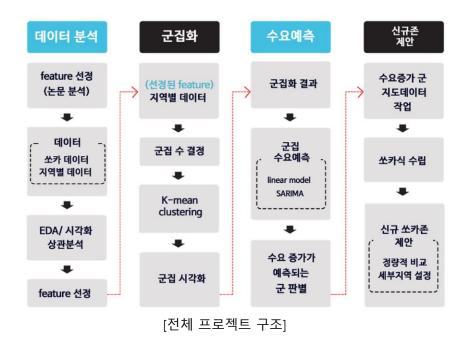
1 A \leftarrow F^T F;
2 Set an integer k sufficiently large enough;
3 Set a nonzero vector x \in \mathbb{R}^m;
4 x_k \leftarrow A^k x;
5 \tilde{\lambda} \leftarrow \frac{x_k^T A x_k}{x_k^T x_k};
6 ||F|| \leftarrow \sqrt{\tilde{\lambda}}
```

Algorithm 2: Approximating ||F|| using the power method

[그림 6-6. power method]

5. 프로젝트 5.2 신규 쏘카존 위치 제안 (AIFFEL)

프로젝트명	신규 쏘카존 위치 제안
프로젝트기간	2022.12~2023.02
프로젝트인원	5명
설명	경기도 지역에 대한 카셰어링 기반 군집화 및 군별 수요예측. 수 요 증가가 예상되는 특정 지역(수원시 팔달구)에 대한 신규 쏘카 존 위치 제안.
담당역할	데이터 취합, 상관분석, 회귀분석, 군집화, 수요예측, 적합도 함수식 수립 등 모든 과정
링크	https://govin08.github.io/machine_learning/socar_zones/
비고	AIFFEL 과정 마지막 프로젝트(AIFFELTHON)





군집화

수요예측

신규존 계안

데이터 분석

소카에서 제공받은 데이터는 지역별, 일자별 수요데이터입니다. 이데이터는 개별적인 소카 사용내역에 대한 기록인데, 주로 경기도권에서 발생한 사용내역들이었습니다. 따라서, 경기도 지역을 타겟으로 삼았습니다. 한편, 수요데이터만으로는 재미있는 결과를 얻기힘들다고 판단해, 지역별 데이터를 만들어냈습니다.

	region1	region2	reservation_return_at	reservation_start_at	age_group	gender	car_model
0	울산광역시	남구	2019-09-29 21:25:40+00:00	2019-09-29 19:20:00+00:00	1	male	경형
1	울산광역시	남구	2019-07-13 13:11:21+00:00	2019-07-13 11:00:00+00:00	1	male	준중형
2	울산광역시	남구	2019-09-16 23:43:08+00:00	2019-09-16 19:10:00+00:00	1	male	준중형
3	울산광역시	남구	2019-08-09 06:17:41+00:00	2019-08-09 00:20:00+00:00	1	male	준중형
4	울산광역시	남구	2019-07-24 12:36:29+00:00	2019-07-24 09:00:00+00:00	2	male	경형

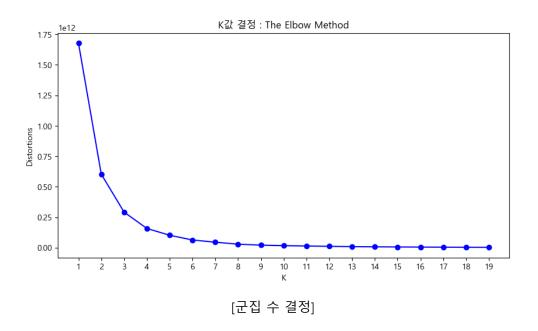
[수요데이터]

	쏘카존 수	인구	2030인구	아파트 수	행복주택 수	지하철역 수	환승역 수	정류장 수	대학교 수	대학생 수	녹지	주거	공업	상업	위도	경도	면적
지역																	
수원시 장안구	18	271012	82294.55980	200	0	1	0	409	0	34441	0.272727	0.727273	0.000000	0.000000	37.303971	127.001743	33.34
수원시 권선구	16	367004	111443.15610	159	1	4	0	537	0	11033	0.625000	0.312500	0.062500	0.000000	37.257687	126.963532	47.17
수원시 팔달구	20	192225	58370.37385	65	1	4	2	185	0	0	0.136364	0.409091	0.000000	0.454545	37.282570	127.011718	12.86
수원시 영통구	34	360127	109354.91020	129	3	6	0	380	0	22911	0.142857	0.571429	0.285714	0.000000	37.259600	127.042341	27.72
부천시	57	791263	224595.00000	520	2	18	2	1110	14	10447	0.230769	0.615385	0.153846	0.000000	37.503592	126.757621	53.44
화성시	56	907958	260316.00000	388	17	3	2	2873	27	14423	0.567568	0.351351	0.054054	0.027027	37.199565	126.823026	693.90

[지역별 데이터]

군집화

지역별 데이터를 통해 군집화(k-means clustering)를 진행했습니다. 이를 통해, 경기도 42개 지역을 여섯 개의 군으로 나누었습니다.

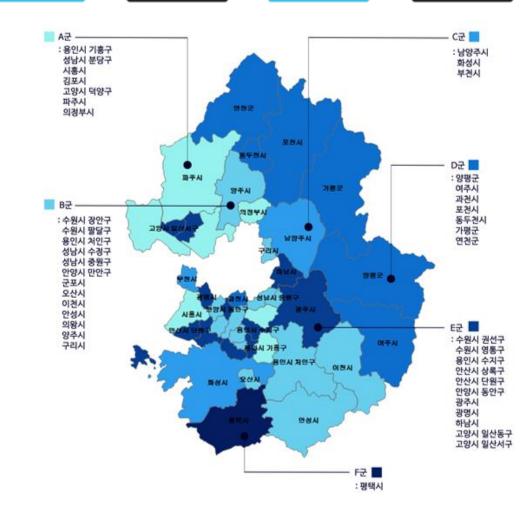


데이터 분석



수요예측





[군집화 결과]



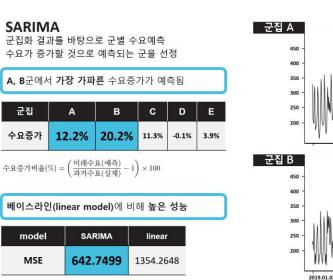
군집화

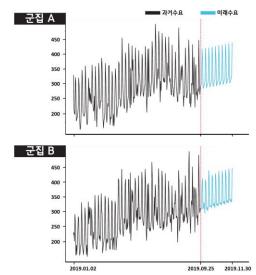


신규존 계안

수요예측

군집화 결과와 수요 데이터를 통해 군별 수요 데이터를 얻어냅니다. 이를 바탕으로 수요예측을 진행해, 미래에 수요가 증가할 것으로 예측되는 군을 선정합니다. 군별 수요 데이터 중 앞의 70% 데이터를 과거 데이터로 뒤의 30%를 미래 데이터로 삼았고, 과거 데이터를 통해 학습하여 미래 데이터를 예측하는 모델(SARIMA)을 만들었습니다. 그 결과, B군이 미래에 수요가 가장 증가할 것으로 예측되었습니다.





[수요예측 결과]

신규 쏘카존 위치 제안

이전 단계에서 경기도의 여러 지역들 중 B군의 수요가 향후에 증가할 것으로 예측되었습니다. 이번에는, B군에서 특정한 한 지역 (수원시 팔달구)을 정해서, 이 지역에 대하여 신규 쏘카존을 제안해봤습니다. 이때, 정성적인 방법 보다는, 정량적인 방법을 동원했습니다.

신규 쏘카존은 공영 주차장에 설치할 것으로 가정했는데, 어떤 공영 주차장이 쏘카존으로 얼마나 적합한지 나타내는 적합도 점수를 생각하고, 이 적합도 점수가 가장 높은 주차장이 쏘카존으로 적절할 것이라고 결론내렸습니다. 적합도 점수는, 교통과 주거, 그리고 쏘카존가지의 거리의 세 개 측면을 고려했고, 추천시스템의 evaluation metric인 DCG에서 아이디어를 얻은 식이 활용되었으며, 정규분포와 일차변환을 통해 적합도 점수가 0과 100 사이의 값이 되도록 설정하여 직접 만들어보았습니다.



군집화

수요예측

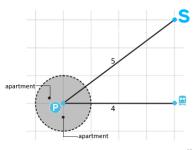




적합도 점수

- 적합도 점수 score = (쏘카존까지의 거리) + (지하철역까지의 거리) + (아파트 밀집도) = score₁ + score₂ + score₃

각 항에 가중치 부여 $score' = a_1 \frac{score_1 - \mu_1}{\sigma_1} + a_2 \frac{score_2 - \mu_2}{\sigma_2} + a_3 \frac{score_3 - \mu_3}{\sigma_3}$ $score_3 = \frac{3}{\sigma_3}$ (반경 ikm 내 세대수)



SCORE: 5 + 4 + 2 = 11



[신규 쏘카존 제안]

5. 프로젝트

5.3 신호생성 및 통행배정 모듈 개발 (Ul networks)

프로젝트명	신호생성 및 통행배정 모듈 개발
프로젝트기간	2023.09 ~ 2024.09
프로젝트인원	2명
설명	성남시 ITS 사업의 일환으로 교통상황 시뮬레이션시 실시간 신호 생성 모듈 및 통행배정 모듈 개발
담당역할	신호생성 모듈 개발, 통행배정 모듈 개발



[교통 시뮬레이션 화면]

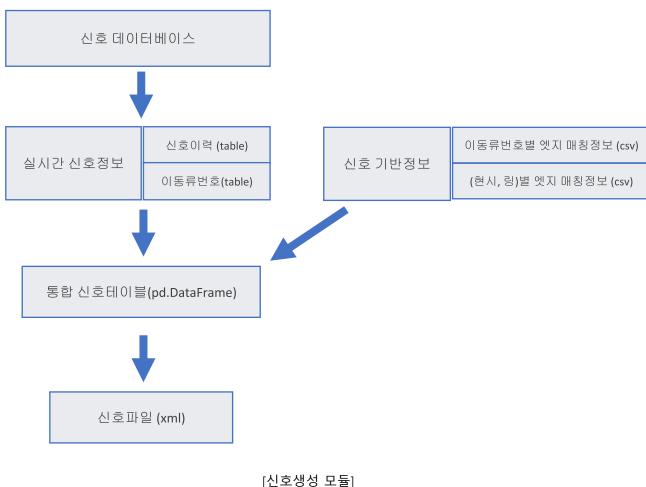
신호생성 모듈

성남시 지능형교통체계 사업 (ITS, intelligent transportation system)은 SKT가 수주한 사업으로, 당사는 해당 사업에 참여하여 교통 시뮬레이션 구현 등에 참여하였습니다.

교통 시뮬레이션은 오픈소스 파이썬 모듈인 sumo(simulation of urban mobility)을 사용해 구현하였습니다. 전년도에도 당사는 sumo를 사용하 여 인천시 ITS 사업에 참여한 바 있습니다.

당사의 기존 신호생성모듈은 실시간으로 전송되는 신호제어기(신호등) 의 신호정보를 반영하지 못했습니다. 단지 신호계획(TOD) 테이블을 사 용하여 이를 통해 일괄적으로 시뮬레이션 신호를 생성하는생성하는 한 계가 있었습니다.

이번에 개발한 신호생성모듈은 신호 DB에 쌓이는 신호 정보를 실시간으 로 수령하여 신호파일 생성에 반영합니다. 또한, 신호정보의 이상치 및 결측치 발생시에도 이를 처리하는 로직이 포함되어 있습니다.

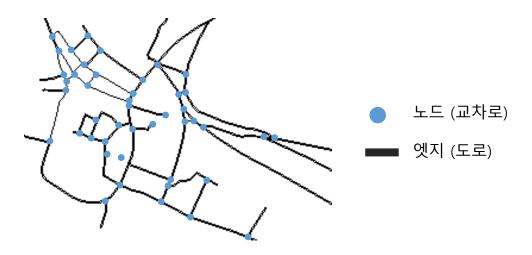


통행배정 모듈

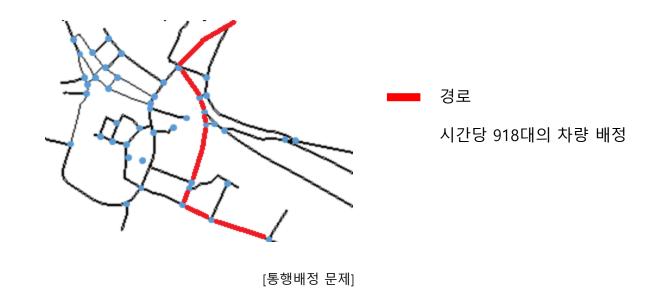
교통 시뮬레이션시 특정 경로에 몇 대의 차량이, 어떠한 분포로 운행하는지를 정해야 합니다. 이것을 교통공학에서는 통행배정 (traffic assignment)이라고 부릅니다.

통행배정을 위해서는 먼저 후보경로를 설정해야 합니다. 이를 위해 교통상황을 directed graph로서 정의하고 (교차로 : 노드, 도로 : 엣지) 파이썬 모듈인 networks를 사용하여 주요 교차로 사이의 최단 경로를 찾는 경로 탐색 알고리즘을 개발했습니다.

후보경로가 설정된 이후에는 해당 경로로 몇 대의 차량이 어떻게 지나갈지에 대해 정해야 합니다. 이를 위해 통행배정 논문 (Sivanandan, 1991)에서 아이디어를 얻어 통행배정문제를 최적화문제로서 정의하고, 선형계획법(linear programming)으로 최적화문제를 풀었습니다. 선형계획법을 구현하는 데에는 scipy 모듈이 사용되었습니다. 차량이 구체적으로 언제 출발(발생)할 지에 대해서는, 차량의 출발이 Poisson process를 따른다고 가정하고, 차량발생간격을 지수분포에서 샘플링하여 구성했습니다. 이상이 통행배정 알고리즘에 대한 설명입니다.



[directed graph로서의 도로 네트워크]



1. 원래 문제

$$\begin{array}{ll} \text{Minimize} & L(x) = M_0 \sum_{(i,j) \in N^2} \sum_{k=1}^{n_{ij}} t_{ij}^k x_{ij}^k + M_1 \sum_{e \in E_*} \left| \bar{f}_e - f_e \right| + M_2 \sum_{(o,d) \in (Z^2)_*} \left| \bar{V}_{od} - V_{od} \right| + M_3 \sum_{\text{inc,out}} \left| \bar{\rho}_{\text{inc,out}} - \rho_{\text{inc,out}} \right| \\ \text{subject to} & x_{ij}^k \geq 0 & \forall i,j,k \\ \\ \text{where} & f_e = \sum_{(i,j) \in N^2} \sum_{k=1}^{n_{ij}} (p_{ij}^k)_e x_{ij}^k & \\ & V_{od} = \sum_{i \in o,j \in d} \sum_{k=1}^{n_{ij}} x_{ij}^k & \\ & \rho_{\text{inc,out}} = \sum_{p_{ij}^k \sim (\text{inc,out})} x_{ij}^k & \\ \end{array}$$

2. 선형계획법 문제로 바꾸기 위해 y, z, w 도입

$$\begin{array}{lll} \text{Minimize} & L(x,y,z,w) = M_0 \sum_{(i,j) \in N^2} \sum_{k=1}^{n_{ij}} t_{ij}^k x_{ij}^k + M_1 \sum_{e \in E_*} (y_e^+ + y_e^-) + M_2 \sum_{(o,d) \in (Z^2)_*} (z_{od}^+ + z_{od}^-) + M_3 \sum_{\text{inc,out}} (w_{\text{inc,out}}^+ + w_{\text{inc,out}}^-) \\ & \text{subject to} & x_{ij}^k \geq 0, \quad y_e^+ \geq 0, \quad y_e^+ \geq 0, \quad z_{od}^- \geq 0, \quad w_{\text{inc,out}}^+ \geq 0, \quad w_{\text{inc,out}}^- \geq 0 \\ & f_e + y_e^+ - y_e^- = \bar{f}_e \\ & V_{od} + z_{od}^+ - z_{od}^- = \bar{V}_{od} \\ & \rho_{\text{inc,out}} + w_{\text{inc,out}}^+ - w_{\text{inc,out}}^- = \bar{\rho}_{\text{inc,out}} \\ & \psi(o,d) \in (Z^2)_* \\ & \psi(\text{inc,out}) \\ & \text{where} & f_e = \sum_{(i,j) \in OD} \sum_{k=1}^{n_{ij}} (p_{ij}^k)_e x_{ij}^k \\ & V_{od} = \sum_{i \in o,j \in d} \sum_{k=1}^{n_{ij}} x_{ij}^k \\ & \rho_{\text{inc,out}} = \sum_{p_{ij}^k \sim (\text{inc,out})} x_{ij}^k \end{array}$$

[통행배정 문제 설명 – 1]

 x_{ij}^k : 교차로 i에서 j로 가는 k번째 경로의 통행량 (vph)

 t_{ij}^{k} : 교차로 i에서 j로 가는 k번째 경로의 통행시간 (s)

 V_{od} : 시작존 o에서 도착존 d로 가는 통행량 (vph)

 f_e : 링크 e에서의 통행량 (vph)

 $\rho_{\text{inc. out}}$: 회전교통량 (vph)

4. 풀어서 표현

5. 행렬식 표현

$$\label{eq:local_equation} \begin{array}{ll} \mbox{Minimize} & L(X) = c^T X \\ \\ \mbox{subject to} & AX = b \end{array}$$

$$A = egin{bmatrix} P & D_E & 0 & 0 \ Q & 0 & D_S & 0 \ U & 0 & 0 & D_H \end{bmatrix}, \qquad b = egin{bmatrix} ar{f} \ ar{V} \ ar{
ho} \end{bmatrix}, \qquad c = egin{bmatrix} M_0T \ M_1 \ M_2 \ M_3 \end{bmatrix}, \qquad X = egin{bmatrix} x \ y \ z \ w \end{bmatrix}$$

[통행배정 문제 설명 – 2]

