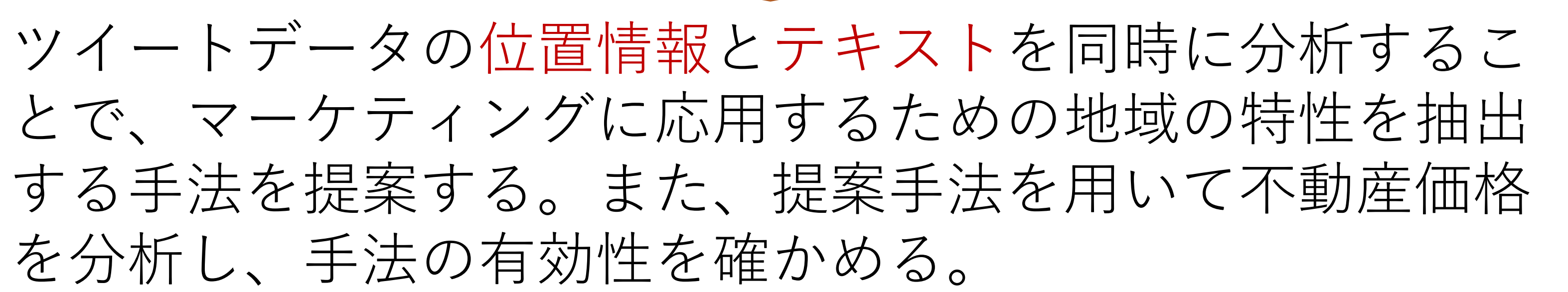


近藤聖也¹ 和泉潔¹ 山田健太¹ 吉田光男²
¹ 東京大学大学院工学系研究科 ² 豊橋技術科学大学
 連絡先: kon.arsenal@gmail.com, izumi-sec@socsim.org

地域の特性を理解することは、マーケティング等の意思決定において重要視されている。またTwitter等のSNSの普及により、人々が自分の活動や感情を発信するようになった。



①ユーザ属性(性別・年代・職業)の推定

$$I_{w,c} = \frac{\frac{1}{n_c} \sum_{j=1}^{n_c} \frac{tweet_count_{j,w}}{tweet_count_j}}{\frac{1}{n} \sum_{i=1}^n \frac{tweet_count_{i,w}}{tweet_count_i}} \quad (1)$$

表1 抽出単語例

性別		年代				職業	
男性	女性	10代	20代	30代	40代	学生	社会人
俺	私	部活	彼	ゴルフ	旦那	学校	酒
オレ	あたし	体育	ライン	巨人	連休	授業	コーヒー
おっさん	プリ	ワロタ	早起き	ビール	(^_^)	課題	ビール
タバコ	ママ	えぐい	チャラ	クラブ	週末	彼氏	会社
会社	ネイル	(。-▽-)	(;_;	店員	自宅	流行り	休み
ギター	♡	ぶす	デート	昭和	研修	テスト	息子
美味しい	かわい	たのしみ	バカ	ノリ	臭い	バイト	娘

表2 ユーザ属性の推定結果

	正解率	ユーザカバー率
性別	88.6%	70.0%
年代	57.9%	76.3%
職業	78.0%	68.0%

※ユーザカバー率＝「不明」を除くサンプル数/全サンプル数

対象地域
山手線全29駅の周辺エリア
(駅を中心として2km×2km
メッシュの地域)

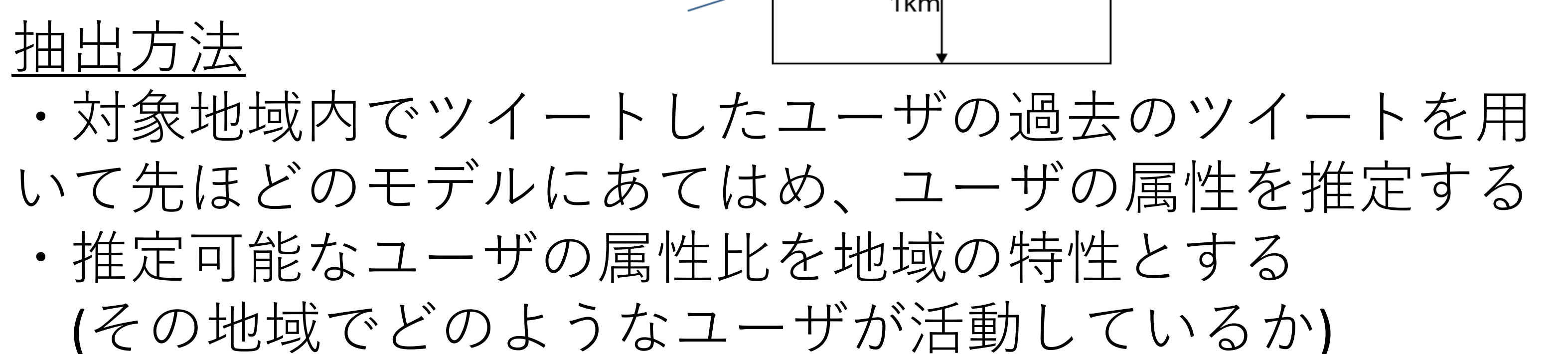
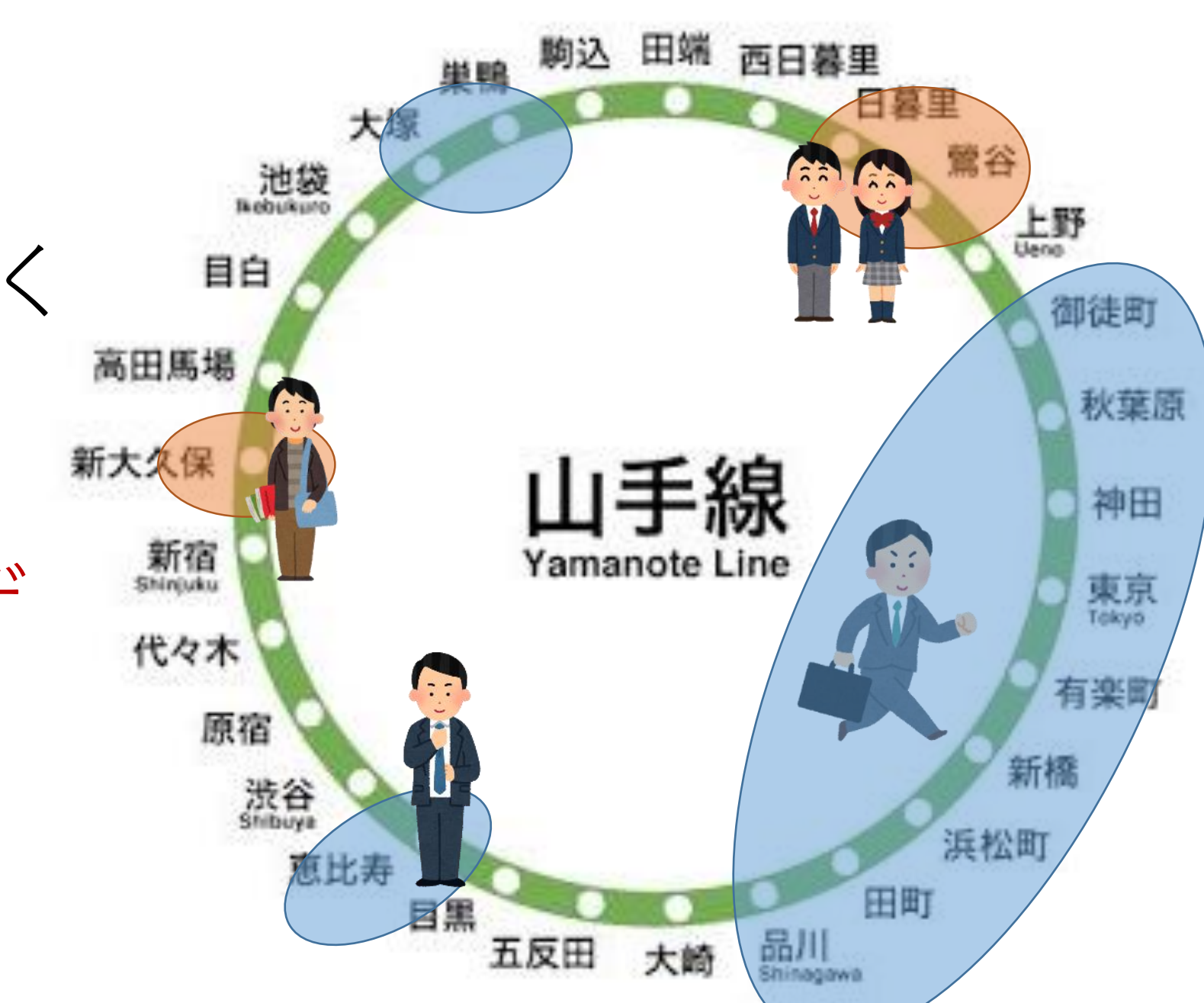


表3 地域クラスタリング結果

番号	1	2	3	4	5
駅	大崎、代々木、新宿、池袋、上野	高田馬場、日暮里、鶯谷	五反田、目白、駒込、田端、西日暮里	東京、有楽町、新橋、浜松町、田町、品川、恵比寿、大塚、巣鴨、御徒町、秋葉原、神田	渋谷、原宿
値の大きい特微量	平日/0時～6時/女性 平日/0時～6時/社会人 休日/6時～12時/女性 休日/12時～18時/20代 休日/18時～24時/学生	休日/12時～18時/学生 平日/18時～24時/学生 平日/18時～24時/10代 平日/6時～12時/学生 休日/0時～6時/男性	休日/18時～24時/40代 休日/6時～12時/学生 平日/0時～6時/学生 休日/6時～12時/40代 平日/0時～6時/40代	休日/0時～6時/社会人 平日/18時～24時/男性 平日/18時～24時/社会人 平日/0時～6時/男性 休日/12時～18時/男性	平日/0時～6時/20代 平日/6時～12時/30代 平日/18時～24時/女性 平日/12時～18時/女性 休日/18時～24時/女性

その地域でどのようなユーザ
が活動しているかが分かる



1. 物件の家賃を予測する回帰分析
被説明変数：家賃
説明変数：面積、駅からの距離、築年数、階数、部屋数
⇒ **決定係数：0.816**

図1 地域ごとの平均残差

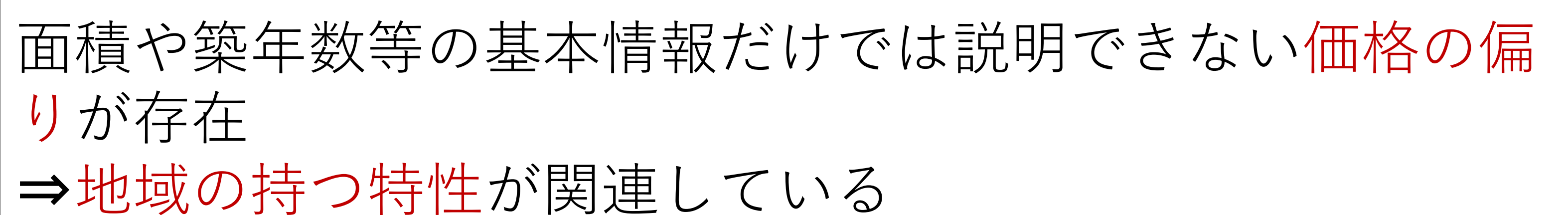


表4 クラスタごとの平均残差

番号	1	2	3	4	5
平均 残差 (円)	-450	-12,253	-9,459	5,108	37,262

2番：学生が多く活動⇒価格が安くなる傾向
4番：男性の社会人が夜に多く活動⇒価格が高くなる傾向

- ・位置情報付きTwitterを用いて地域の特性を抽出する手法を提案した
- ・地域特性を用いて不動産価格を分析した結果、**地域の特性と不動産の価格には関係があることが分かった**

