

特征空间

简体中文版

基础概念 (Foundations) • The Valuation Engineer

Bert Craytor • 2026 年 8 月 1 日

版本 1.0.0

DOI: 10.67330/3yymh010

简体中文译本，由人工智能辅助翻译并经编辑部审校；英文版为权威版本。

形式定义。一类异质商品的特征空间 (characteristics space)，是以买方赋予价值的各项属性为坐标的向量空间；在这一空间中，每件商品由其属性水平构成的向量来表示。

直观阐释。上一条目确立了异质商品是特征的独特组合。自然的下一个问题是：我们应当如何形式化地描述并比较这些组合？答案出自 Lancaster (1966)：把每件商品视为向量空间中的一个点，该空间的坐标轴正是买方在意的各项特征。

在这种表示下，一宗帕西菲卡 (Pacifica) 住宅不再是”帕克帕西菲卡的一栋房子”，而是一个向量，其坐标包括建筑面积、地块面积、景观、状况、学区、距海远近等等。若两宗房地产在所有这些属性上的取值都相同，它们将占据特征空间中的同一个点；在任何属性上存在差异的房地产则占据不同的点。空间的几何结构编码了两宗房地产之间的接近或差异程度，并涵盖了买方作出区分的全部维度。

空间的维数取决于所研究市场中哪些属性是相关的。对典型的住宅估价而言，相关空间可能有十来个维度：物理维度（建筑面积、地块、卧室、卫浴、房龄、质量、状况）、区位维度（学区、景观、距便利设施或不利设施的远近）以及法律 / 经济维度（区划、地役权）。对商业地产估价而言，相关维度转向现金流属性（租约结构、租户质量、出租率），但底层的几何思想是相同的。

特征空间之所以有用，在于它把实务工作者的定性直觉（”这些可比案例与估价对象相似”）转化为定量命题（”这些可比案例在特征空间中距估价对象很近”）。一旦定义了距离，每一个销售比较决策都可以被当作关于空间中位置的陈述来审视。

估价师在何处遇到它。估价师无时无刻不在特征空间中工作，只是通常没有叫出它的名字。销售比较网格的每一行是一条坐标轴；每个可比案例是一个由坐标值构成的行向量；每一项调整都是在保持其他轴不变的情况下沿一条轴移动的尝试。

我们将在本期通篇使用的八个帕西菲卡可比案例直接演示了这一构造。每个可比案例都是由建筑面积、地块面积、景观与状况张成的四维特征空间中的一个点：

可比案例	建筑面积（平方英尺）	地块（平方英尺）	景观	状况	成交价格
A	1,650	9,000	有	良好	\$1,425,000
B	1,650	7,200	无	良好	\$1,280,000
C	1,850	7,500	无	良好	\$1,407,500
D	1,450	7,000	无	良好	\$1,155,000
E	1,700	8,000	有	优秀	\$1,470,000
F	1,550	7,100	无	一般	\$1,177,500
G	1,700	10,500	无	良好	\$1,392,500
H	1,600	7,300	无	良好	\$1,312,500

上一条目中的可比案例 A 与 B 是前两行；其余六个案例对示例作了扩展，我们将在本期余下部分充分利用这些扩展。

每个可比案例都是 $\mathbb{R}^4$ 中的一个点：

- 可比案例 A 位于 (1650, 9000, 1, 2)。
- 可比案例 B 位于 (1650, 7200, 0, 2)。
- 可比案例 G 位于 (1700, 10500, 0, 2)。

景观与状况坐标以数值编码（景观取 $\{0, 1\}$ ，状况取 $\{1, 2, 3\}$ ）；价格不是特征空间的坐标，而是空间中位置的函数。

为何关乎可辩护性。特征空间框架使“可比案例选择究竟是什么”变得更加清晰。可比案例不只是“一宗相似的房地产”，而是特征空间相关区域中靠近估价对象的一个点。可比案例在空间中距估价对象越近，所需的调整越小，得出的指示价值也越可靠。

这一重新表述暴露了非正式网格所掩盖的若干实务问题：

- 哪些维度重要？坐标轴的选择是一个建模决策。只有当市场确实对景观单独定价时，把“景观”设为一条轴才有意义；只有当某个市场细分对“风水朝向”定价时，把它设为一条轴才有意义。可辩护的估价要求为坐标轴本身辩护，而不仅仅是为轴上的取值辩护。
- 适用什么距离度量？两个可比案例的价格可能相差 \$200,000，因为它们在特征空间中相距很远；但距离的计算需要对每条轴进行尺度化。未经变换的欧氏距离把一平方英尺建筑面积等同于一平方英尺地块，而市场并不这样看。
- 可比案例是否位于价格函数表现良好的区域？如果估价对象落在特征空间中采样稀疏的角落，就没有任何可比案例靠近它，价值指示必然是外推性的。

估价实例演算。八个帕西菲卡可比案例占据了四维特征空间中一个相对紧凑的区域。考虑这样一个问题：对于位于 (1,700, 7,500, 1, 2) 的假想估价对象——建筑面积 1,700 平方英尺、地块 7,500 平方英尺、有景观、状况良好——哪个可比案例“最近”？

如果我们不加考虑地使用坐标轴未经尺度化的欧氏距离，地块维度会占据主导地位，因为其数值 (7,000–10,500) 远大于景观 (0–1) 或状况 (1–3)。各案例到估价对象的距离如下：

可比案例	距离（未尺度化欧氏距离）	备注
A	1,501	景观相符；地块占主导
B	308	景观不符；地块接近
C	230	景观不符；按原始距离最近之一
D	559	景观不符；建筑面积偏差
E	502	景观相符；状况不同
F	422	景观与状况均不同
G	3,000	景观不符；地块相距甚远
H	224	景观不符；距离最近

按原始距离，可比案例 H 是最近邻。但可比案例 H 没有景观，而估价对象有。朴素的度量误导了我们：地块维度的数值尺度挤掉了类别型的景观维度，而后者对价格实际上更重要。

一个可辩护的距离度量应当以各条轴的隐含价格 (implicit price) 来尺度化——粗略地说，就是该轴每变动一个单位值多少美元。一旦坐标轴按隐含价格尺度化，景观（约值 \$100,000）便不再被地块（约每平方英尺 \$25）所淹没。于是，在景观上与估价对象相符的可比案例 A 便浮现为恰当的首选可比案例。

这正是后续条目中发展的隐含价格工具所提供的东西：对特征空间的一种有原则的尺度化，使“接近”从几何上的猜测变为源自市场的计算。

参考文献。

- Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2), 132–157.
- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.

交叉引用： 异质商品 (*heterogeneous good*)；特征价格函数 (*hedonic price function*)；隐含价格 (*implicit price*)。

附录：Prolog 形式化

以下机器可读的配套文件 (002-characteristics-space.pl) 在本刊的 Prolog 知识库中对本条目进行了形式化。该文件同时作为可浏览的 Prolog 版本随本文一并发布。

```
%% 002-characteristics-space.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 002
%% Title: Characteristics Space
%% Author: Bert Craytor
%% Concept: characteristics_space
%% Version: 0.1.0-draft

% =====
% DICTIONARY IMPORT
% =====

:- use_module('../tools/dictionary/dictionary').
:- dictionary_release('dictionary-2026.1').

% --- From the dictionary (loaded via :- use_module above) -----
% term(characteristics_space, "Characteristics Space", economic_theory,
% "The characteristics space of a class of heterogeneous goods is
% the vector space whose coordinates index the attributes that
% buyers value, and in which each good is represented by the
% vector of its attribute levels.", published, 'dictionary-2026.1').
%
% term(heterogeneous_good, ...).
% term(hedonic_price_function, ...).
% term(implicit_price, ...).
%
% generalized_by(characteristics_space, lancaster_1966).
% depends_on(characteristics_space, heterogeneous_good).
% -----

:- use_terms([ characteristics_space, heterogeneous_good,
                  hedonic_price_function, implicit_price,
                  comp_selection, sales_comparison_approach,
                  real_estate ]).
:- use_relations([related/2, depends_on/2, generalized_by/2]).

% =====
% ARTICLE SUBJECT
% =====

article_concept(characteristics_space).
article_issue(volume(1), number(1), year(2026)).
```

```

article_id('foundations.2026.002').

% =====
% LOCAL VOCABULARY
% =====
%
% Narrow article-local field names + the dimensions of the worked
% example's characteristics space.

local_term(dimension).
local_term(coordinate).
local_term(euclidean_distance).
local_term(distance_metric).
local_term(nearest_neighbor).

% =====
% FORMAL ATTRIBUTION
% =====

% The construction is due to Lancaster (1966), as noted in citations.
attributed_construction(characteristics_space, lancaster_1966).

% =====
% THE ARTICLE'S CLAIM ABOUT REPRESENTATION
% =====
%
% A heterogeneous good becomes a point in a vector space. Each axis is
% a characteristic buyers value; each good is the vector of its
% attribute levels.

representation_of(heterogeneous_good, point_in_vector_space).

% Examples of dimensions relevant to residential appraisal.
residential_dimension(gla_sqft).
residential_dimension(lot_sqft).
residential_dimension(view).
residential_dimension(cond).
residential_dimension(bedrooms).
residential_dimension(baths).
residential_dimension(age).
residential_dimension(quality).
residential_dimension(school_district).
residential_dimension(distance_to_amenity).
residential_dimension(zoning).

% Examples of dimensions relevant to commercial appraisal — included
% to make the article's point that the dimensions shift with market.
commercial_dimension(lease_structure).
commercial_dimension(tenant_quality).
commercial_dimension(occupancy).
commercial_dimension(cash_flow_profile).

% =====
% APPRAISER ENCOUNTERS
% =====
%
% Mapping appraisal practice onto the characteristics-space framework.

practice_to_space_correspondence(

```

```

    sales_comparison_grid_line, coordinate_axis,
    "A grid line is an axis.>").
practice_to_space_correspondence(
    comp_row, row_vector,
    "A comp is a vector of coordinate values.>").
practice_to_space_correspondence(
    adjustment, movement_along_axis,
    "An adjustment moves along one axis holding others fixed.>").

% =====
% WORKED EXAMPLE
% =====
%
% The 8 Pacifica comps are points in a 4D characteristics space spanned
% by GLA, lot_sqft, view, cond_numeric. Bring in the shared dataset.

:- consult('../..kb/pacifica_comps').

% Axes of the chosen space for this article's analysis.
axis(gla_sqft).
axis(lot_sqft).
axis(view).
axis(cond_numeric).

% Numeric encoding for non-numeric axes.
encode(view, no, 0).
encode(view, yes, 1).
encode(cond_numeric, 1, 1).
encode(cond_numeric, 2, 2).
encode(cond_numeric, 3, 3).

%! coord(?Property, ?Axis, ?Numeric) is nondet.
%
% Numeric coordinate of Property along Axis. Handles symbolic attribute
% values by encoding them.

coord(P, A, N) :-
    attribute_value(P, A, V),
    ( number(V) -> N = V ; encode(A, V, N) ).

%! point(?Property, -Vector) is det.
%
% Vector representation of Property in the 4D characteristics space.

point(P, [GLA, Lot, View, Cond]) :-
    coord(P, gla_sqft, GLA),
    coord(P, lot_sqft, Lot),
    coord(P, view, View),
    coord(P, cond_numeric, Cond).

% =====
% THE HYPOTHETICAL SUBJECT
% =====
%
% The article tests distance metrics against this hypothetical property.

hypothetical_subject(subject).
attribute_value(subject, gla_sqft, 1700).
attribute_value(subject, lot_sqft, 7500).

```

```

attribute_value(subject, view, yes).
attribute_value(subject, cond_numeric, 2).

% =====
% NAIVE EUCLIDEAN DISTANCE (the article's straw distance metric)
% =====
%
% Computes the unscaled Euclidean distance between Property and the
% hypothetical subject. Used to demonstrate that unscaled distances
% mislead: the lot dimension dominates because its numeric range is
% much larger than view (0..1) or cond (1..3).

naive_distance_to_subject(P, D) :-
    point(P, V_P),
    point(subject, V_S),
    sum_of_squares(V_P, V_S, S),
    D is sqrt(S).

sum_of_squares([], [], 0).
sum_of_squares([X|Xs], [Y|Ys], S) :-
    sum_of_squares(Xs, Ys, S0),
    D is X - Y,
    S is S0 + D * D.

% =====
% THE ARTICLE'S CENTRAL OBSERVATION
% =====
%
% Under naive Euclidean distance, Comp H is the nearest neighbor. But
% Comp H has no view and the subject does; view matters more to price
% than the naive metric admits. The fix (developed in entries 003-005)
% is to scale each axis by its implicit price.

claim(naive_metric_misleads,
    "Unscaled Euclidean distance crowds out small-range axes (like view) under large-range axes (
    ↪ like lot_sqft). A market-defensible distance scales each axis by its implicit price." ).

% =====
% DEFENSIBILITY QUESTIONS RAISED
% =====
%
% Three questions the characteristics-space framing makes explicit.

defensibility_question(which_dimensions,
    "The choice of axes is a modeling decision. Defending an appraisal requires defending the axes,
    ↪ not just the values on them." ).
defensibility_question(distance_metric,
    "Untransformed Euclidean distance treats one sqft of GLA as equivalent to one sqft of lot,
    ↪ which the market does not. Distance requires a scaling per axis." ).
defensibility_question(extrapolation_region,
    "If the subject sits in a sparsely-sampled corner of the space, no comp is truly close and the
    ↪ indication is extrapolative." ).

% =====
% INFERENCE RULES
% =====

%! same_point(?A, ?B) is nondet.
%
```

```

% Two properties occupy the same point in the characteristics space.
% (Doesn't occur in the running example, but the article notes that it
% would if all coordinates matched.)

same_point(A, B) :-
    point(A, V),
    point(B, V),
    A \== B.

%! nearest_neighbor_by_naive_distance(-P) is semidet.
%
% The comp closest to the subject under the naive Euclidean metric.
% Used to illustrate the article's point that the naive metric chooses
% Comp H, not Comp A.

nearest_neighbor_by_naive_distance(P) :-
    findall(D-Q,
        ( property(Q), Q \== subject, naive_distance_to_subject(Q, D) ),
        Pairs),
    keysort(Pairs, [_-P|_]).

% =====
% CROSS-REFERENCES
% =====

cross_reference(heterogeneous_good).
cross_reference(hedonic_price_function).
cross_reference(implicit_price).

```