

大規模言語モデル意味距離に基づく加重平均法による多変量解析 (LLM 意味距離加重法)

古橋 武（名古屋大学名誉教授）

2026年1月8日

1 はじめに

多変量解析は、主成分分析や因子分析をはじめとして、社会科学・工学・医学・教育評価など幅広い分野で用いられてきた。これらの手法は、多数の変数に内在する構造を低次元で表現し、データの要約や可視化を可能にする点で強力である。一方で、そこで得られる主成分や因子の意味解釈は、解析結果とは独立に研究者が行う必要があり、その妥当性や再現性が十分に保証されているとは言い難い。

特に、複数の変数を「難度」「有用度」「分かりやすさ」といった概念的な集約語にまとめる操作は、多変量解析の実務において頻繁に行われている。しかし、この集約は多くの場合、単純平均や専門家の経験に基づいて行われており、「なぜその変数群が同一概念に属すると言えるのか」「なぜその重み付けが妥当なのか」といった問い合わせに対して、数理的・客観的な説明を与えることは困難であった。その結果、意味解釈は事後的かつ主観的になりやすく、解析結果の比較や再利用を阻む要因となっている。

この問題の本質は、「意味」と「数値」が異なる次元に存在してきた点にある。従来の多変量解析は数値データの構造を扱うことに長けている一方で、変数や因子が持つ意味的関係性を、解析過程の中で直接扱うことはできなかった。意味的な近さや概念的な類似性は、研究者の暗黙知として前提化されるか、解析後の解釈に委ねられてきたのである。

近年、**大規模言語モデル (Large Language Models; LLMs)** の発展により、自然言語の意味を高次元ベクトルとして表現する意味埋め込み (embedding) が利用可能となった。LLMによる埋め込みは、単語レベルの共起情報にとどまらず、文脈や定義文を含めた意味表現を学習しており、意味的類似性を距離として定量化できる点に特徴がある。この性質は、これまで数理的に扱うことが困難であった「意味」を、多変量解析と接続する可能性を示唆している。

しかし、既存研究の多くは、LLMを用いてクラスタリングや分類、要約といった解析そのものを代替・自動化する方向に焦点を当てており、意味と数値を理論的に橋渡しする枠組みとしての活用は十分に検討されていない。生成AIを解析主体として用いる場合、ブラックボックス性や再現性の問題が生じやすく、従来の統計的枠組みとの整合性も課題となる。

そこで本研究では、生成 AI を解析の代替手段としてではなく、意味的距離を生成する装置として位置づける新たな多変量解析の枠組みを提案する。具体的には、人間が定義した集約概念を定義文として与え、それを LLM の意味空間に写像することで、各変数と集約概念との意味的距離を定量化する。この距離に基づいて変数の重みを決定し、加重平均による集約軸を構成した上で、主成分分析などの古典的多変量解析を適用する。

本手法は、意味設計（人間）、意味距離生成（LLM）、数値解析（統計手法）を明確に分離しつつ統合する点に特徴がある。これにより、変数集約の根拠を定量的に説明可能とし、主成分や因子の意味を事前に設計・検証することが可能となる。本研究では、この枠組みを**大規模言語モデル意味距離に基づく加重平均法による多変量解析**として定式化し、実データを用いてその有効性と解釈可能性を示す。

本論文の構成は以下の通りである。第 2 章では関連研究を概観し、従来の多変量解析および意味処理手法との違いを整理する。第 3 章では提案手法の理論的枠組みと数式定義を示す。第 4 章では技術系セミナーのアンケートデータを対象として、

- (1) 概念設計に基づく 4 概念モデル、
- (2) 概念を再編成した 3 概念モデル、
- (3) 従来の 8 変数による主成分分析

の 3 手法を比較し、意味距離に基づく概念集約が主成分構造および解釈性に与える影響を検討する。最後に第 5 章で、3 手法を比較し、意味距離に基づく概念集約が主成分構造および解釈性に与える影響を議論する。

2 関連研究

2.1 心理測定学と尺度構成

因子分析は、多数の観測変数に潜在する少数の共通因子を仮定し、相関構造を説明する枠組みとして発展してきた。能力検査の相関構造から一般因子を導入した Spearman (1904) [1] に始まり、複数因子を想定する多因子分析が Thurstone (1931) [2] によって体系化された。その後、探索的因子分析 (EFA) はデータ駆動で因子構造を探索する手法として、確認的因子分析 (CFA) は仮説構造を明示し、適合度を検証する手法として整理され、構造方程式モデリング (SEM) の測定モデルとして確立された (Jöreskog, 1969 [3]; Brown, 2006 [4])。

尺度構成 (scale development) の実務では、概念定義、項目作成、項目分析、信頼性・妥当性検証という段階的手続きが強調され、因子分析は項目の次元性確認や尺度精錬の中心的役割を担ってきた (DeVellis, 2016 [5])。また、信頼性指標として Cronbach の α が広く用いられてきたが (Cronbach, 1951 [6]), τ 等価性などの仮定が現実のデータでは成立しにくいことが指摘され、McDonald の ω や SEM に基づく信頼性推定法が代替手法として提案されている (Okada, 2011 [7])。

しかし、これらの手法はいずれも、**項目がどの概念を測定しているか、どの項目がより中心的か**といった意味論的判断を、理論や専門家判断として事前に与えることを前提としており、意味設計そのものは統計モデルの外部に置かれてきた。すなわち、数値的には整合的であっても、「なぜこの項目群がこの概念を構成すると言えるのか」という問い合わせに対して、再現可能な数理的説明を与えることは困難であった。

2.2 前処理としての項目集約

実務上、アンケート項目を「難度」「有用度」「学習性」などの上位概念にまとめ、合成得点として扱う操作は広く行われている。しかし、その多くは単純平均や合計、あるいは経験的な重み付けによって実装されており、重みの根拠は専門家判断に依存しやすい。結果として、集約軸は利便性が高い一方で、再現可能な規則として明示しにくいという問題を抱えている。

尺度構成の教科書においても、概念定義と項目内容の整合性 (content validity) は重視されるが、意味的近さそのものを距離として定量化し、集約規則に直接組み込む数学的装置は、基本的に前提化されるに留まっている (DeVellis, 2016 [5])。

2.3 テキスト埋め込みと意味距離

自然言語処理分野では、語や文書の意味をベクトルとして表現し、距離や類似度に基づいて比較する研究が長く行われてきた。共起行列の低ランク近似に基づく Latent Semantic Analysis (LSA) (Deerwester et al., 1990 [8]) や、トピック分布を確率モデルとして導入する Latent Dirichlet Allocation (LDA) (Blei et al., 2003 [9]) がその代表例である。

その後、ニューラルネットワークに基づく分散表現として word2vec (Mikolov et al., 2013 [10]) や GloVe (Pennington et al., 2014 [11]) が提案され、語彙の意味関係をベクトル演算として扱う枠組みが普及した。近年では Transformer 系モデルに基づく文脈依存埋め込みが主流となり、BERT (Devlin et al., 2019 [12]) や Sentence-BERT (Reimers & Gurevych, 2019 [13]) により、文レベルの意味距離を高精度に扱うことが可能となっている。

2.4 埋め込みと心理測定の接続

近年、質問紙項目文を埋め込み化し、項目間の意味的類似性を用いて因子構造の事前推定や尺度短縮を行う研究が報告されている。これらの研究では、回答データに依存せずに項目の意味構造を分析できる点が利点とされ、尺度設計や事前検証への応用が示されている。

さらに、文埋め込みを用いた項目削減 (scale shortening) や、項目・尺度・ラベルを同一意味空間に配置して分類体系の歪み (jingle / jangle fallacy) を検出・整理する研究も提案されている。しかし、これらの研究の多くは、意味空間そのものを解析対象とする立場に立ち、数値データの集約や多変量解析の前処理として意味距離を組み込む枠組みには至っていない。

2.5 本研究の位置づけ

以上を踏まえると、既存研究は大きく二つの系譜に整理できる。一方は、因子分析や SEM に代表される数値データの構造検証に強い枠組みであり、意味設計は外生的的前提として扱われる。他方は、埋め込みを用いた意味解析であり、意味構造の定量化には優れるが、数値データの前処理や統計解析との接続は限定的である。

本研究は、大規模言語モデルを推論主体として用いるのではなく、**意味距離生成器**として位置づけ、人間が定義した集約概念と各項目との意味的距離を重みへ変換し、項目群ごとの重み付き平均として数値データを集約する。その上で、主成分分析や因子分析を探索手法ではなく、意味設計の妥当性を検証する手段として用いる点に特徴がある。

すなわち、本研究は、「意味を解析するために LLM を用いる」のではなく、「意味を数値解析に組み込むために LLM を用いる」という立場を明確にし、尺度構成とテキスト埋め込みを橋渡しする実務的かつ再現可能な方法論を提示するものである。

3 提案手法

3.1 提案手法の概要

本研究では、大規模言語モデル（LLM）の意味空間を用いて変数と概念の意味的距離を定量化し、その距離に基づいて数値データを重み付きで集約した上で、多変量解析を適用する枠組みを提案する。本手法を**大規模言語モデル意味距離に基づく加重平均法による多変量解析**と呼ぶ。

本手法は以下の 4 段階から構成される。

1. 集約概念の定義と変数グループの設計
2. LLM による意味埋め込み
3. 概念意味距離に基づく重み付けと数値集約
4. 集約後データに対する多変量解析

本研究では、概念ごとに定義された変数グループ内で意味距離に基づく重み付けを行う方法（グループ × 意味距離集約）を提案手法として採用する。

3.2 記号定義と問題設定

観測データを以下のように定義する。

- n : サンプルサイズ
- p : 元変数の数
- $\mathbf{X} = (x_{ij}) \in \mathbb{R}^{n \times p}$: 数値データ行列
- X_j : 第 j 変数 ($j = 1, \dots, p$)

集約概念集合を

$$\mathcal{C} = \{C_1, C_2, \dots, C_K\}$$

とし、各概念は自然言語による定義文として与えられる。

3.3 大規模言語モデルによる意味埋め込み

各変数および集約概念に対し、対応する説明文・定義文を用意し、大規模言語モデルを用いて意味埋め込みを行う。

$$\mathbf{x}_j = f_{\text{LLM}}(T_j), \quad \mathbf{c}_k = f_{\text{LLM}}(D_k) \quad (1)$$

ここで、

- T_j : 変数 X_j に対応する自然言語による説明文
- D_k : 集約概念 C_k を定義する自然言語文

また。 $\mathbf{x}_j, \mathbf{c}_k \in \mathbb{R}^d$ は意味ベクトルであり、距離計算のために正規化される。 d は大規模言語モデルが生成する意味埋め込みベクトルの次元数である。

3.4 概念意味距離と重み定義

変数 X_j と概念 C_k の意味距離をコサイン距離で定義する。

$$d_{jk} = 1 - \frac{\mathbf{x}_j^\top \mathbf{c}_k}{\|\mathbf{x}_j\| \|\mathbf{c}_k\|} \quad (2)$$

各概念 C_k に対応する変数インデックス集合を

$$\mathcal{J}_k \subseteq \{1, 2, \dots, p\}$$

として与える。 \mathcal{J}_k は概念 C_k に関連付けられた変数群を表す。

\mathcal{J}_k は事前知識に基づき与えられる場合に加え、意味距離に基づく閾値処理やクラスタリングによって自動的に決定される場合にも対応可能である。

概念 C_k に属する変数群内で、以下の重みを定義する。

$$w_{jk} = \frac{(d_{jk} + \varepsilon)^{-1}}{\sum_{l \in \mathcal{J}_k} (d_{lk} + \varepsilon)^{-1}}, \quad j \in \mathcal{J}_k \quad (3)$$

ここで ε は数値安定性のための十分小さい正定数である。

このとき、

$$\sum_{j \in \mathcal{J}_k} w_{jk} = 1 \quad (4)$$

が成り立つ。

この定義は、各概念ごとに独立した正規化を行うものであり、意味設計に基づく局所的な集約を重視している。

3.5 集約データの構成

各変数は事前に標準化される。

$$\tilde{X}_j = \frac{X_j - \mu_j}{\sigma_j} \quad (5)$$

ここで μ_j および σ_j はそれぞれ変数 X_j の標本平均および標本標準偏差である。

概念 C_k に対する集約データを次式で構成する。

$$Z_k = \sum_{j \in \mathcal{J}_k} w_{jk} \tilde{X}_j \quad (6)$$

3.6 集約データに対する多変量解析

集約データ $Z \in \mathbb{R}^{n \times K}$ に対して、例えば主成分分析 (PCA) を適用し、評価構造の低次元表現を得る。

3.7 方法論上の仮定

本研究の枠組みは、質問項目および概念定義文が、大規模言語モデル (LLM) によって安定的かつ一貫して意味ベクトルへ写像され、その間の距離が人間の直感的な意味的近接性を十分に近似しているという仮定に基づいている。また、得られた意味距離を重みとして用いた加重平均が、元の数値データに含まれる情報構造を本質的に歪めることなく、概念レベルでの集約を可能にすることを前提としている。さらに、本研究では、概念文の定義を解析者が事前に与えるという立場を採っており、この概念設計が分析目的に対して妥当であることを仮定している。

4 実験 (Experiment)

4.1 用いたデータ

表 1: アンケート調査データ

回答者ID	内容の難度	内容の有用度	製作課題の難度	製作課題の有用度	レポート課題の難度	レポート課題の有用度	ビデオ、資料の見易さ	穴埋め方式は集中しやすい形式	全体満足度
1	3	5	3	5	3	5	5	5	5
2	4	4	2	4	4	3	3	4	2
3	3	5	3	5	3	5	4	4	5
4	4	5	3	5	4	4	4	3	4
5	5	4	3	4	5	4	2	4	2
6	4	5	4	4	4	4	4	3	4
7	2	4	1	3	3	4	4	4	3
8	2	4	3	4	4	4	2	3	3

本研究では、技術系セミナーの受講者を対象として実施されたアンケート調査データを分析対象とした。表1は、得られたアンケート結果を示す。アンケートは、講義内容、教材・提示手法、実践課題およびレポート課題に関する評価項目から構成されており、各項目は Likert 型尺度により数値化されている。各尺度において、数値が大きいほど、当該評価内容への肯定的な度合いが高いことを意味する。

分析に用いたデータの概要を以下に示す。

- サンプルサイズ：8名
- 評価項目数：9項目
 - 主分析対象項目：
内容の難度、内容の有用度、製作課題の難度、製作課題の有用度、レポート課題の難度、レポート課題の有用度、ビデオ・資料の見易さ、穴埋め方式は集中し易い形式
 - 補助変数：全体満足度

回答者IDおよび全体満足度は、主成分分析の入力変数からは除外し、可視化および解釈補助のためにのみ使用した。

4.2 実験条件

本研究では、同一データに対して以下の3手法を適用し、結果を比較した。

4.2.1 手法1：4概念モデル（提案法）

提案手法 LLM-SSWMA を用い、以下の4概念を定義した。

- **講義内容の専門的適合性**：セミナー講義内容が受講者の専門的背景および期待水準に対して適切であったかを評価する。
- **実践課題の設計妥当性**：製作課題の難度と有用度から実践的学習設計の妥当性を評価する。
- **思考深化課題の設計妥当性**：レポート課題が理解深化および内省的学習に寄与したかを評価する。
- **教材・提示手法の理解支援性**：教材および提示形式が理解および集中をどの程度支援したかを評価する。

各概念に対応する質問項目グループは、以下のように定めた。

- **講義内容の専門的適合性**：“内容の難度”、“内容の有用度”
- **実践課題の設計妥当性**：“製作課題の難度”、“製作課題の有用度”

- 思考深化課題の設計妥当性：“レポート課題の難度”、“レポート課題の有用度”
- 教材・提示手法の理解支援性：“ビデオ・資料の見易さ”、“穴埋め方式は集中しやすい形式”

4.2.2 手法2：3概念モデル（提案法・概念再設計）

同一の提案手法を用い、概念設計のみを変更し、以下の3概念を定義した。

- **難度**：学習者が内容を理解し、課題を遂行する際に要求される認知的負荷や困難さの程度を表す概念
- **有用度**：学習内容や課題が、学習者にとってどの程度価値があり、実践的・知的に役立つと認識されているかを表す概念
- **設計品質**：教材や課題の構成、提示方法、形式が、学習を支援する観点からどの程度適切に設計されているかを表す概念

各概念に対応する質問項目グループは、以下のように定めた。

- **難度**：“内容の難度”、“製作課題の難度”、“レポート課題の難度”、“ビデオ・資料の見易さ”
- **有用度**：“内容の有用度”、“製作課題の有用度”、“レポート課題の有用度”、“穴埋め方式は集中しやすい”
- **設計品質**：“ビデオ・資料の見易さ”、“穴埋め方式は集中しやすい形式”

ここで、「難度」のグループには“ビデオ・資料の見易さ”を、「有用度」のグループには“穴埋め方式は集中しやすい形式”を、それぞれ意図的に含めている。本構成は、意味的に必ずしも自然とは言えない質問項目をあえて含めることで、概念意味距離の定量化手法が概念定義と質問項目との意味的整合性の違いをどの程度識別できるかを検証することを目的としたものである。

4.2.3 手法3：従来の主成分分析

比較対象として、8個の質問項目をそのまま変数とし、標準化後に通常の主成分分析を適用した。

4.3 実験手順

すべての手法について、以下の手順を共通とした。

1. 回答データの標準化
2. （提案法のみ）概念意味距離に基づく集約データの構成
3. 主成分分析の適用
4. 主成分得点・負荷量・全体満足度との関係の可視化

4.4 実験結果

4.4.1 手法 1：4 概念モデル

(1) 概念意味距離

表 2: 概念と概念内質問項目との距離

概念	質問項目	距離
講義内容の専門的適合性	内容の難度	0.578
	内容の有用度	0.559
実践課題の設計妥当性	製作課題の難度	0.377
	製作課題の有用度	0.258
思考深化課題の設計妥当性	レポート課題の難度	0.429
	レポート課題の有用度	0.368
教材・提示手法の理解支援性	ビデオ・資料の見易さ	0.607
	穴埋め方式は集中し易い形式	0.627

各概念グループに属する質問項目と、対応する概念定義文との意味距離を表 2 に示す。ここで示した距離は、式 (2) に基づくコサイン距離であり、値が小さいほど質問項目文と概念定義文の意味的類似性が高いことを表す。

いずれの概念においても、対応づけられた質問項目は概念定義文とかけ離れた距離（1.0 に近い値）にないことが分かる。

講義内容の専門的適合性 「内容の難度」と「内容の有用度」はほぼ同程度の距離を示しており、両項目が概念を均等に代表していることが確認された。

実践課題の設計妥当性 「製作課題の有用度」は「製作課題の難度」よりも小さな距離を示しており、概念定義文との意味的近接性が高い。

思考深化課題の設計妥当性 「レポート課題の有用度」は「難度」よりもやや小さな距離を示している。

教材・提示手法の理解支援性 「ビデオ・資料の見易さ」と「穴埋め方式は集中し易い形式」は、いずれもほぼ同程度の距離を示している。

以上の結果から、意味距離に基づく分析は、単に重み付けを行うためだけでなく、各質問項目が概念定義をどの程度意味的に代表しているかを定量的に検証する手段としても有効であることが示された。本データにおいては距離の差は限定的であったが、これは概念設計と質問項目設計が意味的に一貫していたことを裏付ける結果であり、意味距離を用いた検証の妥当性を示すものと位置づけられる。

(2) 主成分負荷量に基づく主成分軸の解釈

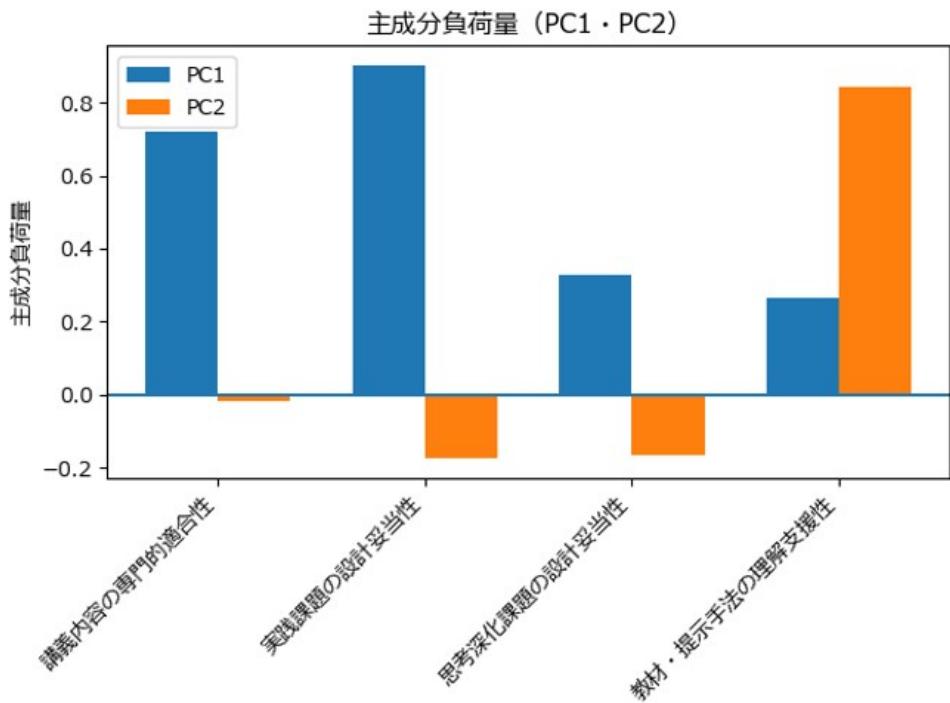


図 1: 4 概念モデルの主成分負荷量

4 概念モデルにおける主成分負荷量を図 1 に示す。橙色が第 1 主成分 (PC1)、青色が第 2 主成分 (PC2) の負荷量である。横軸は左から「講義内容の専門的適合性」、「実践課題の設計妥当性」、「思考深化課題の設計妥当性」、「教材・提示手法の理解支援性」である。

第 1 主成分 (PC1) 第 1 主成分 (PC1) は、「講義内容の専門的適合性」および「実践課題の設計妥当性」に対して特に大きな正の負荷量を示した。一方で、「思考深化課題の設計妥当性」および「教材・提示手法の理解支援性」の寄与は相対的に小さい。

この負荷構造から、PC1 は、

セミナーの中核的価値（内容の適合性と実践課題の妥当性）を表す軸

として解釈できる。

すなわち PC1 の正方向は、「専門的に適切で、実践課題を通じて学習価値が高い」と評価した受講者を表し、負方向は、その逆の評価傾向を示す。

第 2 主成分 (PC2) 第 2 主成分 (PC2) は、「教材・提示手法の理解支援性」に対して強い正の負荷を示す一方、「実践課題の設計妥当性」および「思考深化課題の設計妥当性」に対して負の負荷を示した。

のことから PC2 は、

学習体験の質的差（理解支援性と課題妥当性の対比）を表す軸

として解釈できる。

すなわち PC2 の正方向は、「理解支援性は高いが、課題の妥当性はやや低い」評価を、負方向は、「課題妥当性はやや高いが、理解支援性が低い」評価を反映している。

(3) 主成分得点分布の特徴とその解釈

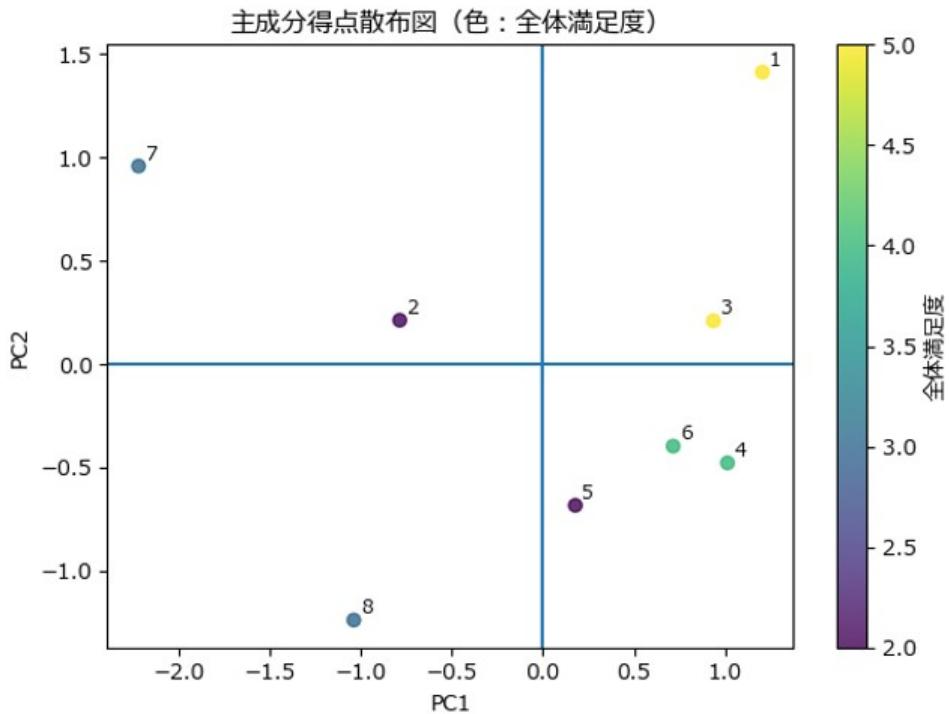


図 2: 4 概念モデルの主成分得点

上記の主成分軸の意味づけを踏まえ、主成分得点散布図を解釈する。図2に4概念モデルの主成分得点の散布図を示す。横軸は第1主成分（PC1）、縦軸は第2主成分（PC2）であり、各点は各受講者の主成分得点を表している。各点の右上の数字は受講者IDである。また、点の色は全体満足度を表し、黄色である（数値が大きい）ほど満足度が高いことを示す。

PC1 方向の特徴（中核的価値） 主成分得点散布図において、全体満足度の高い受講者は PC1 の正方向に集中して分布している。これは、

セミナーの満足度が、内容の専門的適合性および実践的価値と強く結びついていることを示している。

一方、PC1 の負方向に位置する受講者は、内容や実践課題に対する評価が相対的に低く、満足度も低い傾向を示している。

PC2 方向の特徴（学習体験の質的差） PC2 方向では、全体満足度が同程度であっても、受講者が正負に分離する様子が観察された。これは、

満足度が高い受講者の中にも、「分かりやすさ重視型」と「課題妥当性重視型」という異なる評価様式が存在することを示唆している。

すなわち、

- PC2 正方向：
教材・提示手法の理解支援性を高く評価し、比較的負担の少ない学習体験
- PC2 負方向：
製作・レポート課題の負荷は高いが、思考深化や実践性を重視した学習体験
という 学習体験の質的差が明確に分離されている。

4.5 手法2：3概念モデル（概念再設計）

(1) 概念意味距離

表 3: 再設計概念と質問項目との意味距離

概念	質問項目	距離
難度	内容の難度	0.448
	製作課題の難度	0.593
	レポート課題の難度	0.549
	ビデオ・資料の見易さ	0.729
有用度	内容の有用度	0.443
	製作課題の有用度	0.442
	レポート課題の有用度	0.443
	穴埋め方式は集中し易い形式	0.805
設計品質	ビデオ・資料の見易さ	0.629
	穴埋め方式は集中し易い形式	0.736

再編成した3概念（難度・有用度・設計品質）と、対応づけられた質問項目との意味距離を表3に示す。

難度 “内容の難度”、“製作課題の難度”、“レポート課題の難度”は、いずれも比較的近い距離を示している。一方、“ビデオ・資料の見易さ”は他の項目と比べて大きな距離を示しており、難度概念との意味的近接性が相対的に低いことが確認された。

有用度 “内容の有用度”、“製作課題の有用度”、“レポート課題の有用度”はほぼ同程度の距離を示している。一方、“穴埋め方式は集中し易い形式”は距離が大きく、有用度という概念との意味的近接性が相対的に低いことが分かる。

設計品質 “ビデオ・資料の見易さ”と“穴埋め方式は集中し易い形式”は、いずれも中程度の距離を示している。その中で、“ビデオ・資料の見易さ”の方が「設計品質」により近い位置にあることが分かる。

なお、意図的に含めた質問項目（「難度」概念に対する“ビデオ・資料の見易さ”，「有用度」概念に対する“穴埋め方式は集中しやすい形式”）はいずれも、当該概念との意味

的近接性が低い値を示しており、概念間の識別性が意味距離によって適切に反映されていることが確認された。

(2) 主成分負荷量に基づく主成分軸の解釈

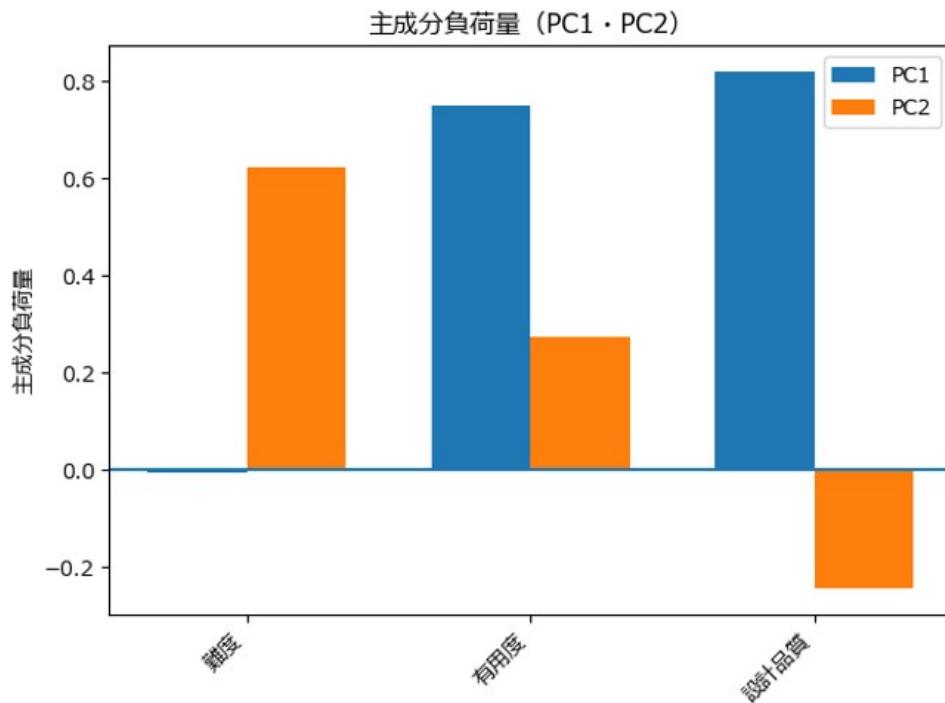


図 3: 3 概念モデルの主成分負荷量

3 概念モデルにおける主成分負荷量を図 3 に示す。第 1 主成分 (PC1) および第 2 主成分 (PC2) の負荷構造は以下の特徴を示した。

第 1 主成分 (PC1) 「有用度」および「設計品質」に対して大きな正の負荷量を示しており、一方で「難度」の寄与はほとんど見られない。このことから、PC1 はセミナーに対する価値評価や設計上の品質を反映する総合的有用性・品質軸として解釈できる。

第 2 主成分 (PC2) 「難度」に対して強い正の負荷量を示し、「設計品質」に対しては負の負荷量を示している。一方、「有用度」は中程度の正の寄与を持つ。この負荷構造から、PC2 は学習負荷や困難さの程度を表す学習負荷軸として解釈され、設計品質との対比関係を内包した軸であると考えられる。

(3) 主成分得点分布の特徴とその解釈

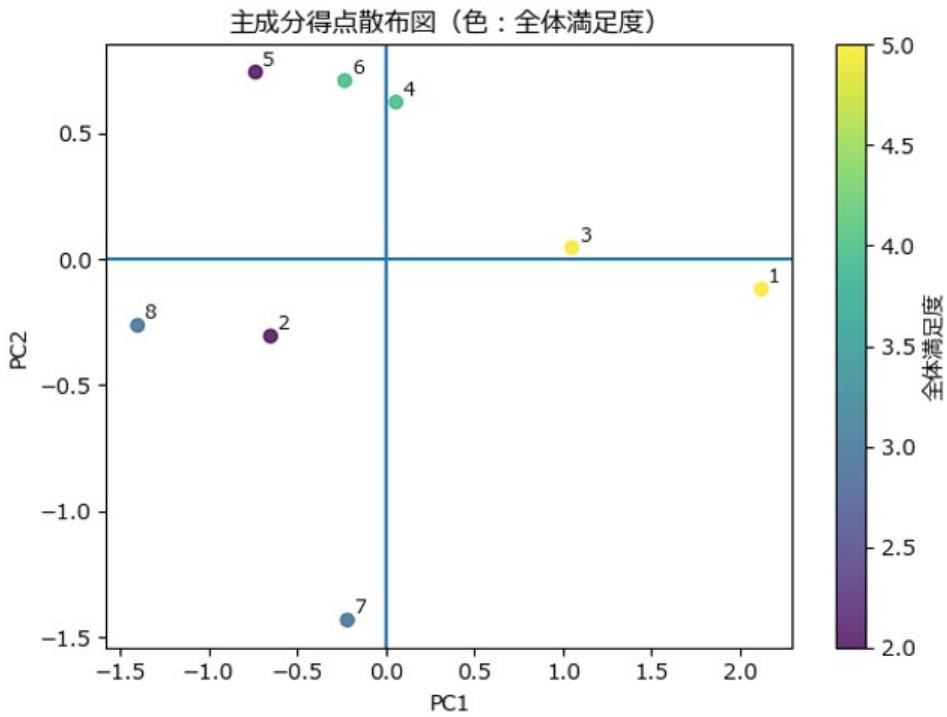


図 4: 3 概念モデルの主成分得点

図 4 は、3 概念（難度・有用度・設計品質）に対して意味距離集約を行った後の主成分得点散布図を示している。点は各受講者を表し、色は全体満足度を示す。

PC1 の方向 PC1 の正方向には、全体満足度の高い受講者が集中して分布しており、セミナーの満足度が有用度や設計品質と強く関連していることが確認できる。一方、PC1 の負方向には満足度の低い受講者が位置しており、セミナーに対する総合的価値評価の低さが満足度の低下と対応していることが示唆される。

PC2 の方向 PC2 の正負方向に着目すると、受講者 4 および 6 は、難度が比較的高いと評価しているにもかかわらず、全体満足度が高いことが分かる。これは、学習負荷が高くても、学習内容の価値や設計品質が十分であれば、高い満足度が得られる可能性を示している。

また、同程度の全体満足度を示す受講者であっても、PC2 方向に沿って分布が大きく分かれている様子が観察される。これは、セミナーを低く評価した受講者の中にも、「負荷が高いにもかかわらず学習価値が低い」と評価する者と、「負荷が低く、かつ学習価値も低い」と評価する者が存在することを示している。

以上より、本モデルでは、PC1 によってセミナーの総合的価値が明確に抽出され、PC2 によって学習負荷に関する評価の違いが補助的に表現されている。この二軸構造により、全体満足度と学習体験の質的差異を同時に可視化できている点が、本分析の特徴である。

4.6 手法3：従来の主成分分析（8変数）

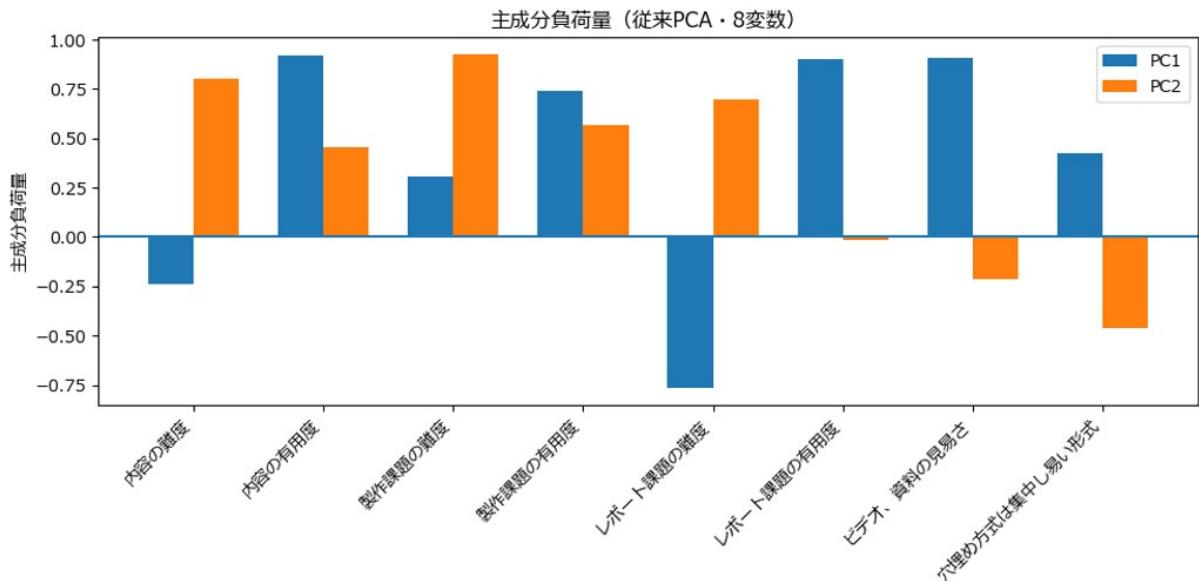


図 5: 従来法による主成分負荷量

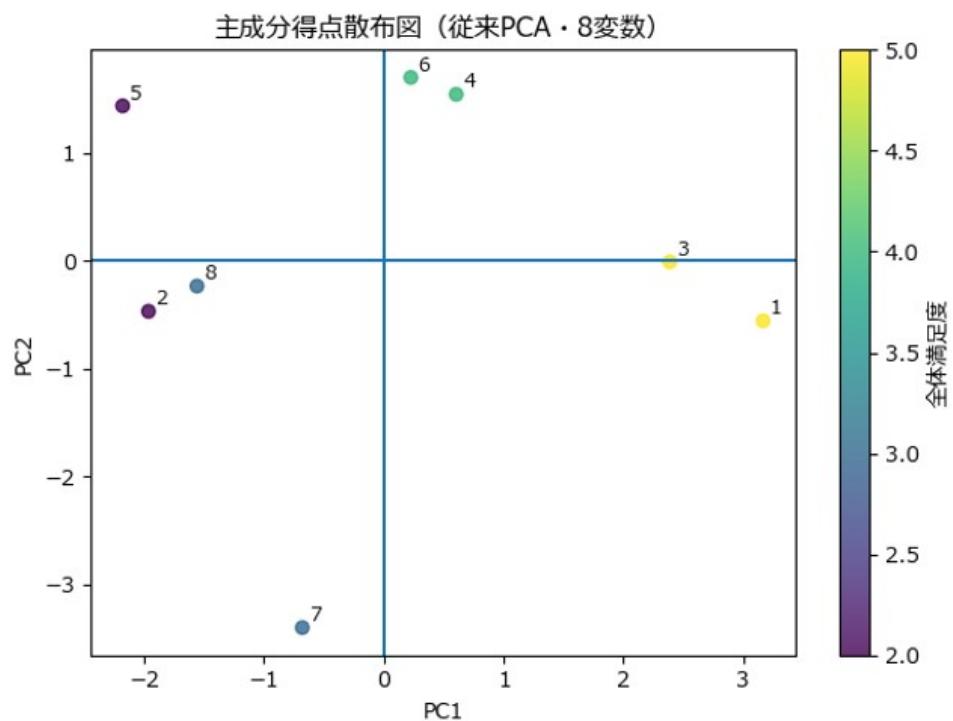


図 6: 従来法による主成分得点

図 5 および図 6 は、8変数を用いた従来法の主成分分析結果によって得られた主成分負荷量と主成分得点を示している。

(1) 主成分負荷量に基づく主成分軸の解釈

第1主成分（PC1） 内容・課題・レポートに関する“有用度”項目および教材・提示手法に関する評価項目が正の負荷量を示す一方，“難度”に関する項目が負の負荷量を示す構造が確認された。この結果、有用度評価と難度評価という意味的に異なる評価次元が、同一の主成分上に混在した軸が形成されている。

第2主成分（PC2） 講義内容や課題に関する“難度”項目が主として正の負荷量を示しており、学習時に受講者が感じた認知的・作業的負荷を反映する軸として解釈可能である。

(2) 主成分得点の特徴とその解釈

主成分得点散布図においては、

PC1の方向 PC1の正方向に全体満足度の高い受講者が多く分布し、PC1が満足度と強く対応する総合評価軸であることが確認された。

PC2の方向 一方で、同程度の全体満足度を示す受講者であっても、PC2方向には大きな分散が見られ、学習負荷の感じ方には個人差が存在することが示唆される。

以上の解釈は、3概念モデルにおける主成分構造と本質的に類似しており、従来法においても、「総合的価値評価」と「学習負荷」という二つの側面が主成分空間上で抽出されていることが分かる。

5 考察

本研究では、(1) 概念設計に基づく4概念モデル、(2) 概念を再編成した3概念モデル、(3) 従来の8変数を用いた主成分分析の3手法を比較し、意味距離に基づく概念集約が主成分構造および解釈性に与える影響を検討した。

5.1 主成分構造に共通する特徴

いずれの手法においても、第1主成分（PC1）は全体満足度と強く対応しており、セミナーに対する総合的な価値評価を表す軸として抽出された。また、第2主成分（PC2）は、学習負荷や評価様式の違いを表す補助的な軸として機能していた。

のことから、本データにおいては、「総合的価値評価」と「学習負荷」という二つの主要な評価側面が、手法によらず安定して存在していることが確認された。

5.2 従来法と3概念モデルの関係

従来の主成分分析（8変数）と3概念モデルの結果は、主成分構造および主成分得点の分布において本質的に類似していた。両手法ともに、PC1が満足度と強く対応し、PC2が学習負荷の個人差を反映する構造を示している。

しかし、従来法では、有用度・難度・設計に関する項目が主成分上で混在しており、主成分負荷量の解釈には一定の恣意性が残る。これに対し3概念モデルでは、概念単位で変数が整理されているため、主成分軸の意味づけがより明確になっている。

すなわち、3概念モデルは、従来法に内在していた評価構造を、意味的に整理された形で再表現したモデルであると位置づけられる。

5.3 4概念モデルにおける差異の顕在化

4概念モデルでは、従来法および3概念モデルと比較して、主成分構造の解釈性がさらに向上した。

第1主成分では、「講義内容の専門的適合性」と「実践課題の設計妥当性」が強く寄与し、セミナーの中核的価値が明確に抽出された。第2主成分では、「教材・提示手法の理解支援性」と課題系概念との対比構造が明瞭に現れ、学習体験の質的差異が可視化された。

この結果、同程度の全体満足度を示す受講者であっても、「分かりやすさ重視型」と「課題妥当性・思考深化重視型」といった評価様式の違いが、主成分空間上で明確に分離された。これは、概念を細分化したことにより、評価構造の分解能が高まった結果と考えられる。

5.4 意味距離に基づく概念集約の意義

意味距離分析の結果、4概念モデルでは、各質問項目が概念定義と意味的に整合していることが確認された。距離の差自体は大きくなかったが、これは質問項目設計と概念設計が意味的に一貫していたことを示す結果である。

一方、3概念モデルでは、意図的に意味的に不自然な項目を含めた結果、対応する概念との距離が大きくなり、意味距離が概念間の識別性を反映することが確認された。

のことから、意味距離は単なる重み付け手法にとどまらず、概念設計の妥当性を検証し、概念間の混入や不整合を検出する指標として有効であることが示唆される。

5.5 総合的評価

以上より、従来の主成分分析は評価構造の大枠を把握する上では有効であるが、解釈の明確性には限界があることが分かった。3概念モデルは、従来法の結果を意味的に整理し、解釈性を高める役割を果たしている。さらに4概念モデルでは、評価構造の分解能が向上し、学習体験の質的差異まで可視化できる点で、最も実務的有用性が高いと考えられる。

本研究は、意味距離に基づく概念設計と数値解析の統合が、アンケートデータ解析における解釈性と設計改善への示唆を同時に高め得ることを示した。

5.6 限界

一方で、本手法は概念定義文の記述内容や表現方法に一定程度依存しており、概念文の書き換えや使用する LLM・埋め込みモデルの違いが意味距離や集約結果に与える影響については、本研究では体系的に検証していない。また、意味距離に基づく重み付けが数値データの分散構造や統計的性質に与える影響については理論的な保証が限定的であり、本手法は従来の主成分分析や因子分析を代替するものではない。加えて、本研究で用いたデータ規模や項目数は限定的であり、より大規模・多様なデータセットに対する一般性については今後の検討課題である。

6 結論

本研究では、大規模言語モデルの意味空間を用いて、質問項目と評価概念との意味距離を定量化し、その距離に基づいてアンケートデータを概念単位に集約した上で主成分分析を行う手法を提案した。さらに、4概念モデル、3概念モデル、および従来の8変数による主成分分析を比較し、意味距離に基づく概念設計の有効性を検証した。

その結果、いずれの手法においても、セミナーに対する総合的価値評価と学習負荷という二つの主要な評価側面が主成分空間上に抽出された。一方で、概念設計の粒度によって、評価構造の解釈性と分解能に顕著な差が生じることが確認された。

特に、4概念モデルでは、概念設計に基づく集約によって主成分軸の意味が明確化され、同程度の満足度を示す受講者間における学習体験の質的差異が可視化された。また、意味距離分析により、質問項目と概念定義との整合性を定量的に検証できることが示された。

以上より、本研究は、大規模言語モデルを解析主体として用いるのではなく、意味距離を一貫して生成する前処理装置として明確に位置づけることで、概念定義に依存するという仮定と限界を自覚的に受け入れつつも、多数の変数を概念レベルで整理し、従来の数値中心の多変量解析では困難であった解釈可能性と構造的洞察を同時に高める、実用的かつ再現可能な分析枠組みを提示した。本手法は、教育評価アンケートをはじめとする、意味的解釈が重要な調査データ分析に対して、有効な分析枠組みを提供するものと期待される。

参考文献

- [1] Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, 15(2), 201–293.
- [2] Thurstone, L. L. (1931). Multiple Factor Analysis. *Psychological Review*, 38(5), 406–427.
- [3] Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34(2), 183–202.
- [4] Brown, T. A. (2006). Confirmatory Factor Analysis for Applied Research. *Guilford Press*.

- [5] DeVellis, R. F. (2016). Scale Development: Theory and Applications. Sage.
- [6] Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), 297–334.
- [7] 岡田謙介 (2011). クロンバックの α に代わる信頼性の推定法について. 心理学研究.
- [8] Deerwester, S., et al. (1990). Indexing by latent semantic analysis. Journal of the American Society for Information Science, 41(6), 391–407.
- [9] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3, 993–1022.
- [10] Mikolov, T., et al. (2013). Efficient estimation of word representations in vector space. arXiv:1301.3781v3.
- [11] Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. Proc. of Conf. on Empirical Methods in Natural Language Processing. 1532–1543.
- [12] Devlin, J., et al. (2019). BERT: Pre-training of deep bidirectional transformers. Proc. of NAACL-HLT. 4171–4186.
- [13] Reimers, N., & Gurevych, I. (2019). Sentence-BERT. arXiv:1908.10084v1.