

金融工学の数学のための 知識データベース構築の提案

白田由香利* 橋本隆子† 久保山哲二‡

要旨

本研究の目的は、金融工学数学教育のための学習支援システムの構築である。難度の高い金融関連文章題が与えられたときに、学生が演繹推論することを、支援するシステムを構築したい。この学習支援システムは、学生が解にいたるプロセスを演繹推論することを支援する。システムは知識ベースを検索し、必要な公式を探し、それを使った演繹推論の最適なパスを提示する。技術的課題としては、用語のユニークなセマンティクスのラベル付け、文章題日本語文からの Given Data と Unknown の抽出、公式のメタレベルでの記述、最適な解法プランの作成、等がある。

Proposal of Knowledge Database Construction for Mathematical Finance

Yukari Shirota* Takako Hashimoto† Tetsuji Kuboyama‡

Abstract

The purpose of this research is construction of a computer-assisted learning system for financial mathematics. Given a hard bond mathematics word problem, the system helps a student conduct deductive reasoning to solve the problem. The system retrieves the knowledge database to obtain the formula that is required for the reasoning so that the system may construct and offer the solution path to the unknown. The problems we should overcome include systematic semantics labeling of bond mathematics terminology, extraction of Given Data list and Unknown from the given word problem, description of formula stored in the knowledge database, and construction of the optimal solution plan.

* 学習院大学 経済学部経営学科, Gakushuin University, Department of Management, Faculty of Economics

† 千葉商科大学 商経学部, Chiba University of Commerce, Faculty of Commerce and Economics

‡ 学習院大学 計算機センター, Gakushuin University, Computer Center

1. はじめに

本稿では、金融工学の数学教育のための学習支援システムの構築を提案する。

経済学部で数学を教えてきた経験から感じることは、多くの経済学部学生が経済数学を難しいと考えていることである。その理由として、(1)経済に関する知識及び、数学に関する知識の不足、(2)知識からの演繹推論力の不足、があげられる。経済数学の場合、数学の知識と、経済に関する知識が複雑に交じり合うため、演繹推論プロセスが複雑になるので、問題解法のための演繹推論が正しく行えないのである。本研究で扱う推論の種類は演繹推論のみであるので、以下、単に推論と言った場合は演繹推論のこととする。さらに、学生の問題解決を困難にさせる理由としては、業界用語の不統一性による問題の影響も大きい。

経済数学教育のため、白田は、知識ベースに基づく推論エンジンシミュレーション法(以下、推論エンジン法と略す)という数学教授法を創案し、実際の講義においてこの教授法を実践してきた[1-3]。推論エンジン法の特長は、学生に2種類の知識ベースを、頭の中に構築させる点である。経済の知識ベースには、経済の公式を、数学の知識ベースには数学公式を覚えさせる。学生アンケートの結果からも、この教授法の効果が高いことは示された[1]。

我々は、推論エンジン法に基づき、学生が数学文章題を解こうとする際、それを支援する知識ベース・システムの構築を行いたい。学生は、解にいたるプロセス(以下、解法プランと呼ぶ)を演繹推論して求めようとするが、それが困難な際に、システムが正しく演繹推論するための知識を提示し、演繹推論の最適なパスを提示することで、学生を支援する。

広い意味での我々の研究対象は、汎用の数学文章題であるが、知識ベースは対象とするドメインによって、非常に異なってくるので、研究ターゲットドメインの特定が必要となる。そこで、ターゲットドメインとして、我々は、「債券ポートフォリオの管理」という分野を選んだ。この分野の数学文章題は、現代の金融機関の債券およびポートフォリオを扱うので、その重要性は年々増えている。しかしながら、その問題解法プロセスを理解するためには、数学と債券に関する2種類の知識を必要とするため、理解できる学生の数は限られてくると言えよう。

数学の知識としては、微分、偏微分、及びテイラー展開に関する公式が必要であり、債券の知識としては、複利計算法、各種デューレーション、イミュニゼーションに関する公式が必要となる。本提案システムの支援によって、債券ポートフォリオの管理の数学文章題を理解する学生数が増えることを目標とする。

我々は、債券ポートフォリオの管理に関する数学文章題を対象とする、学習支援シ

システムの構築を提案する。本システムの入力文章は日本語文章である。システムからの第1番目の出力は、文章題で与えられたデータ(以下、Given Data と呼ぶ)及び、文章題で聞いている未知数(以下、Unknown と呼ぶ)である。例として、以下のような文章題が与えられたとする(引用元資料[4]):

ある債券(最終利回り 4%)のマコーレー・デュレーションは 5 年である。最終利回りが 1% 下落したときの、この債券の価格変化率を計算しなさい。

この文章は、原文のままである。図 1 では、この問題を推論エンジン法で解くプロセスを示した。

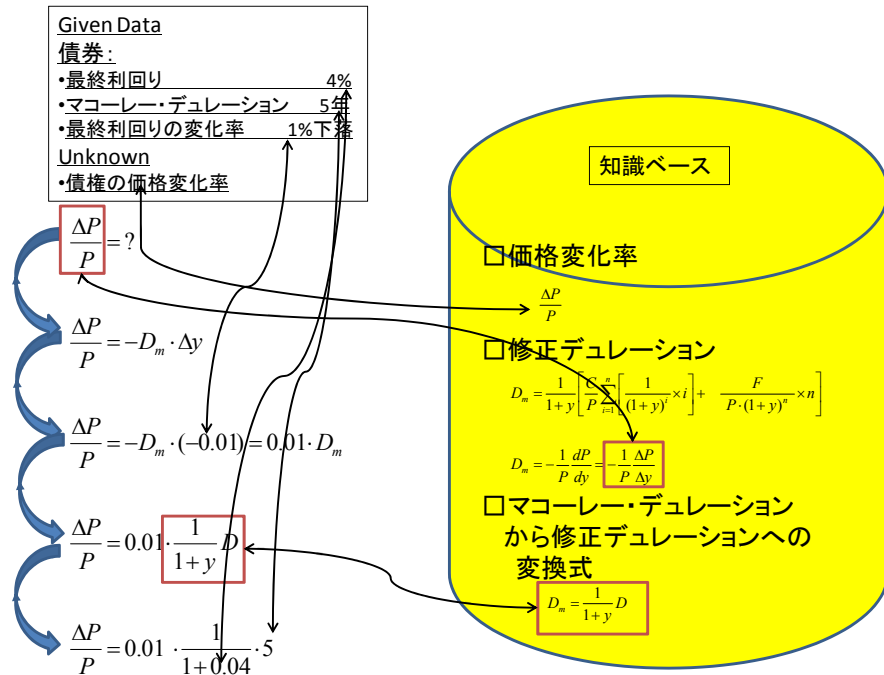


図 1:文章題から Given Data 及び Unknown を抽出し、知識ベースから必要な公式を検索し、演繹推論を行う。

図中右上に Given Data のリストと Unknown が記されている。学生が Given Data と Unknown を抽出できない場合は、学生の読解スキルを補完するため、システムが Given data 及び、Unknown を抽出し、提示する。

学生もシステムも推論エンジン法に基づいて問題を解いていく。得られる解法プランは一つとは限らないが、得られる解法プランは人がやってもシステムがやっても、意味的には同じ結果となる。解法プランとは、公式を含む推論プロセスのパスである。

未知数の値を推論するためには、問題解決に必要な公式の検索が必要である。学生が知識不足のため、必要な公式が分からない、あるいは、正確に思い出せない場合には、システムが代わって、知識ベースから必要な公式を検索して提示する。

図 1 の例で説明する。まず、Unknown は債券の価格変化率である。キーワード「価格変化率」を検索すると、その定義式 $\frac{\Delta P}{P}$ が発見できた。次に、この定義式 $\frac{\Delta P}{P}$ の値を計算するために、この定義式を含む公式を検索する必要がある。検索すると、修正デュレーションの公式が発見できた。修正デュレーションの式は 2 種類あるが、 $\frac{\Delta P}{P}$ を含んでいるのは 2 番目の式である(図 1 参照)。

修正デュレーションの式の変形から、価格変化率の式 $-D_m \cdot \Delta y$ が求められる。y が最終利回りの変数であることは、知識ベースの公式に記されているので、分かるとする。同様に、 Δy が最終利回りの変化率であることが分かる。

次に、キーワード「最終利回りの変化率」を求めて Given Data リストを検索すると、検索はヒットして、1%の下落、つまり、 $\Delta y = -0.01$ であることが分かる。

次に、価格変化率の計算を進めるために必要な値は、 D_m (修正デュレーション)である。しかし、修正デュレーションの値は Given Data には無い。しかし、マコーレー・デュレーション関連の公式の中には発見できた。そこで、その中のひとつ、 D (マコーレー・デュレーション)から修正デュレーションへの変換式の公式 $D_m = \frac{1}{1+y} D$ を使

うことにする。この最適パスを取得するプロセスにおいては、解に到達しないパスを刈る必要もある。

現在の最終利回り y 及び、マコーレー・デュレーションの値は文章題で与えられている: $y=0.04, D=5$ 。よってこれらの公式を使った演繹推論の結果として、価格変化率の値が求まる。

本研究の目的は、こうした金融工学数学教育のための学習支援システムの構築である。難度の高い金融関連文章題が与えられたときに、学生が演繹推論することを、支

援するシステムを構築したい。この学習支援システムは、学生が解にいたるプロセスを演繹推論することを支援する。システムは知識ベースを検索し、必要な公式を探し、それを使った演繹推論の最適なパスを提示する。以下では、第2節でシステム構成を論じる。第3節では、既存研究との関連を述べる。第4節はまとめとする。

2. システム構成

本節では、システムについて説明する。図2は提案システムの構成図である。本システムはユーザインタフェース部、言語処理部、推論エンジン部(含む数式処理部)、提示情報生成部の4つの処理部と、コーパス、知識ベースの2つのデータベースから構成される。各処理部について説明する。

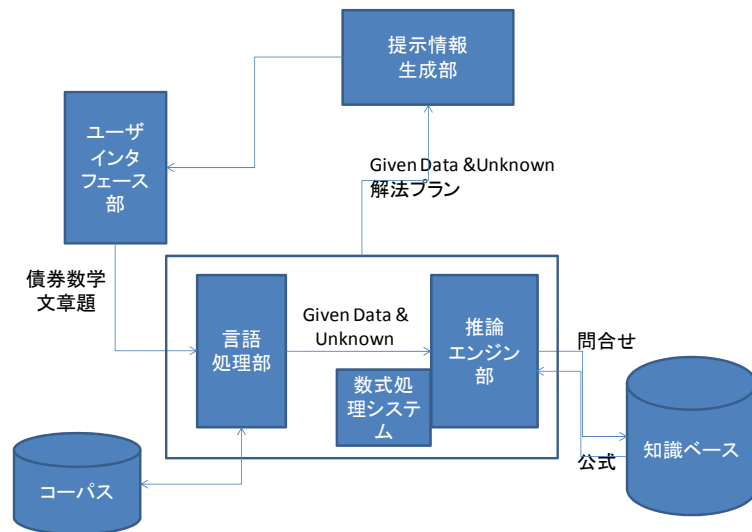


図2:システム構成図

- ユーザインタフェース部

本システムの利用者である学生は本処理部を通じて、数学の問題(文章題)を入力し、解法についての支援を受ける。数学文章題は教師から与えられる場合や、学生自らが問題集から自分の解きたい文章題を設定する場合などがあるが、いずれにせよ数学の文章題の日本語文章が、入力データとなる。今回のシステムでは、対象分野が債

券ポートフォリオ管理に関する数学文章題であるので、債券数学の文章題が次の処理部に渡されることになる。

- 言語処理部

本処理部はユーザインタフェース部から送られてきた債券数学文章題が、入力データとなる。言語処理部では、言語処理プロセスにより、Given Data 及び Unknown が抽出される。抽出にあたっては、あらかじめ用意されているドメイン(今回の場合は債券ポートフォリオ管理)に応じたコーパスを利用する。

債券数学文章題は省略が多く、比較的短い文章であり、係り受けも正確に記載されていないことが多い。暗黙の了解で、記載されていないデータも多く存在する。例えば、債券の額面は100であることが、慣例である。そうした文章を対象に、ドメインのコーパスを活用しながら効率的に Given Data と Unknown を抽出するところが、本システムの一つのキーポイントとなる。例として挙げた文章題の場合、Given Data と Unknown は以下のように抽出される。

Given Data

債券

最終利回り 4%
マコーレー・デュレーション 5年
最終利回りの変化率 1%下落

Unknown

債券

価格変化率

上記の抽出結果を見てもわかるように、Given Data や Unknown は階層化されている。債券というオブジェクトには複数の属性があるので、文章題で問われているのはどのようなオブジェクトか、その属性のうち何が Given Data となり、Unknown となるかを的確に判定する必要がある。例えば、ポートフォリオ作成に関する文章題では、債券 A,B,C のように複数の債券が出てくるため、どの債券に関する属性であるのか、明示する必要がある。

- コーパス

本システムでは、専門用語ごとに、その品詞と、ユニークに振られたセマンティックスのラベル付けを行う。例えば、「イミュニゼーション:名詞:イミュニゼーション」、というようにラベルを付ける。

債券の世界は歴史が古いため、業界用語が複雑で、用語のゆらぎが大きいことが問

題である。業界内のグループによっても使用する用語が異なる。これが初心者にとって、大きな障壁となる。例えば、債券価格の文章題中で、「利回り」、「最終利回り」、「満期利回り」、「yield」などが、同じ概念「満期利回り」を表わすために使われている。

著者は、現実の金利の影響を受けて変動する場合は「利回り」、期間中一定値をとる場合は「最終利回り」と用語を使い分けるようにしているが、こうしたルールは書き手に依存して異なる。これが債券数学文章題の複雑さを増幅してしまう。その用語のゆらぎを取るための、一意のセマンティクスのラベリングが必要であり、それが本システムの特長となるであろう。

もちろん文脈に依存して、ひとつの用語に複数のセマンティクスが対応する場合も

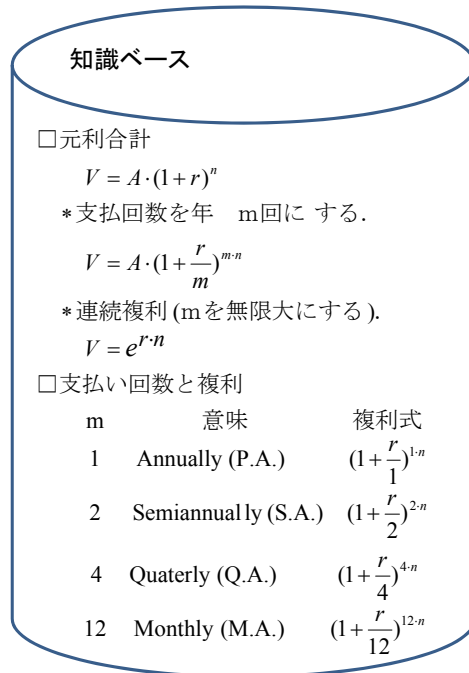


図 3: 支払い回数と複利に関する公式

ある。例えば、「デュレーション」という用語に対して、多くの場合は、その概念は、「修正デュレーション」であるが、債券のイミュニゼーションに関する問題においては、「デュレーション」は、「マコーレー・デュレーション」を意味する。この文脈依存を、Given Data だけから判別可能かどうか、研究していく必要がある。

● 知識ベース

知識ベースには、数学の公式と、経済の公式が含まれる。例えば、等比数列の級数の公式が数学の公式であり、価格感度の公式は経済の公式である。

債券数学の公式を知識ベースに格納する際の最大の問題は、公式表記のゆらぎの問題である。この問題は、債券数学に特有のものであるので以下に説明する。

[債券数学における公式表記のゆらぎ]

● 流派によって、同じ概念を表わす記号が異なる。

例えば、Luenberger の Investment Science[5, 6]では、利回りは記号 y を使うが、他の多くのテキストでは、 λ を使う。一般的な数学問題であるのだから、定義して使えば、記号として何を選択しても問題はないのであるが、身体(視覚)の反応による理解の速さを考えると、学生によっては、表記の違いは内容理解の障壁となる。

● 複利計算法の公式として、支払い頻度により、以下の3つのモードがある。

- 支払い頻度は1年に1回 (Per Annually)
- 支払い頻度は1年にm回 (半年ごとであれば $m=2$, 月ごとであれば $m=12$ 等)
- 連続複利. 上記2で、 $m \rightarrow \infty$ とした場合. 公式に、ネピアの e の指数が含まれるようになる。

これらを教える際には、図3にある公式のように学生に説明している。この3つのモードは、他の債券数学の公式の多くに適用されるので、殆ど全ての債券数学の公式を3つのモードで表記する必要がある。例えば、将来価値(Future Value)や修正デュレーションの公式である。

債券数学の文章題が、数学としてはそれほど難しくはないにもかかわらず解けない理由として、上記2つの公式表記のゆらぎの問題、及び、用語の不統一性の問題がある、と我々は考える。公式表記のゆらぎの問題に関しては、記号を、どのテキストの書き方に準拠するのか、及び、複利の支払い回数をいくつに設定するのか、これらを、学生がモード選択可能にする必要があると、我々は考える。

例えば、学生が「Luenberger 流表記で、半年複利」と指定したならば、そのように記号の選択と公式の変形がなされるべきである。そのためには、より上位の、セマンティクスレベルでの公式表記が必要である。また、指定モードによって式を変換して

する機能が必要である。

各種のモード毎に総当り的に組み合わせのケースを作るような、また、場当たりに膨大な数のケース分けした公式を格納するのは、知識ベースの悪い作り方である。このような知識ベースは、学生の手本とはならない。我々は、債券数学の分野で、学生の手本となるような公式の表記方式を確立したい。体系づけられたメタレベルの公式表記は、エレガントな知識ベースを構築に役立つであろう。

● 推論エンジン部

本処理部は言語処理部のアウトプットである **Given Data** と **Unknown** から、その問題を解くために利用する公式を推論する処理部である。公式が蓄積されている知識ベースを検索することで、推論を実施していく。

第1節であげた例題の場合、**Unknown** の価格変化率という単語から、その定義式 $\frac{\Delta P}{P}$

を検索結果として得る。さらにこの、この定義式 $\frac{\Delta P}{P}$ を解くために、この定義式を含む公式を検索する。すると、修正デュレーションの2番目の定義式が発見できる。人間の場合は、式の変形を行いながら推論を柔軟に行うが、システムの場合、式の変形の実現は容易ではないので、数式処理システムを呼び出して式の変形を行う必要がある。

式の検索を行うための、オンデマンドな数式変換、及び、セマンティクスレベルのラベルによる検索が必要となる。これらも研究対象となる。式の部分一致を判別する機能を実現するためには、上述したユニークなセマンティクスのラベル付けが必要である。

途中、公式をノードとするパスは複数に枝分かれするので、その中から **Unknown** に至るパスを選別する必要がある。与えられた文章題に対して、最適かつ最短な解法プランを導き出し、適切な公式を検索する部分が本研究のもう一つのキーポイントとなる。ここで言う、解法プランとは、公式を組み合わせることで演繹推論する過程パスのことである。この解法プランが提示できれば、学生の学習効率は大きく向上するであろう。

● 提示情報生成部

本処理部は、言語処理部のアウトプットである **Given Data** と **Unknown**、推論エンジン部が導き出した解法プランに沿った解法ステップを、学生に提示する部分である。特に解法プランのステップ毎の提示については、学生とのインタラクションによるフィードバック機能を加味し、見せ過ぎず、学生に考えさせるように、適度な開示度で示していくことが望ましい。この効果的な提示方法も本研究のポイントとなる。

3. 関連研究

数学教育において生徒・学生の文章問題理解を支援するシステムは数多く提案されている[7]。数学の文章問題を解決するためには、言葉の理解能力と問題分野の数学の能力、および、両者を総合し、文章を対象領域の文脈から捉えて問題解決のプランを立てる能力が必要となる[8]。

例えば、**MathCAL**[9]は算数教育において、ポリヤの問題解決手続きに基づき、(1)問題理解、(2)プラン設計、(3)プラン実行、(4)解答の評価、の4段階において生徒を支援するシステムである。小学5年生レベルの問題を対象に、**MathCAL**を用いて算数学習をする群と用いないで学習する対照群との間で、実際に学習前と学習後の算数のテストの成績を比較することで効果測定を行い、有効性を示している。**MathCAL**はあらかじめ保持した問題と解決手順を提示してゆくシステムであり、本研究のように未知の問題には対応していない。

ROBUST[10]は算数の文章問題を解く際の「理解のプロセス」についてある認知モデルを仮定し、実際にそのプロセスをシミュレートするシステムである。複数の算数の問題解決が、システムにより系統的に解けることを示すことで、提案認知モデルの有効性を検証している。とくに、文章中の単語列のパターンを算数の演算概念にマッピングする過程を場合分けして、さまざまな文章問題に対応できるようにしているのが特徴である。

しかし、**ROBUST**は人の認知科学的な問題理解の過程を実際に検証している訳ではなく、算数の文章問題が実際にシステムで解けたか否かを評価の手掛かりにしているため、「理解のプロセス」と謳っている対象は **STUDENT**[11]等の古典的な算数の文章問題解決システムを緻密化して、より幅広い算数の文章問題が解けるようにパターンマッチングをチューニングしたに過ぎないともいえる。

LIM-G[12]は初等的な幾何学の文章問題に対して、解答のための手掛かりを提示するシステムである。本研究と同様に、あらかじめ問題とその解答戦略を対で保持せず、柔軟に未知の問題に対応するための知識ベースを備えている。しかし、基本的に1段階の推論のみ支援し扱える幾何学の問題領域も狭い。

4. まとめ

金融工学数学教育のための文章題の解法のための学習支援システムの構築を本稿で提案した。対象ドメインとする債券ポートフォリオの管理は学生にとって難解な分野である。こうした難度の高い金融関連文章題が与えられたときに、学生が演繹推論することを、支援するシステムの構築が主たる研究目的である。

債券の分野は歴史が古いため、用語の統一性がないため、用語のゆらぎ、公式表記

のゆらぎは大きな問題となる。そのためには、ユニークなセマンティックレベルでのラベル付けをしたコーパス作成が必要であり、また、公式の数式表記にも、メタレベルな表記が必要となる。我々はこうした特有の問題を解決していきたい。

技術的課題としては、用語のユニークなセマンティクスのラベル付け、文章題日本語文からの Given Data と Unknown の抽出、公式のメタレベルでの記述、最適な解法プランの作成、等がある。

48(4): p. 582-601.

謝辞

本研究の一部は、平成 22 年度科研費基礎研究(C)一般「推論エンジン法による知識ベースの構築」(課題番号: 22500231)による、ここに記して謝意を表します。

参考文献

1. 白田由香利, 推論エンジンをベースとした経営数学教授法とそのアンケート評価. 学習院大学経済論集, 2010. Vol.46(No.3).
2. 白田由香利, 悩める学生のための経済・経営数学入門 3つの解法テクニックで数学アレルギーを克服!. 2009, 東京: 共立出版.
3. 白田由香利, 推論エンジンをベースとした経営数学教授法. 情報処理学会研究報告, 情報学基礎(FI), 2009. Vol.2009-FI-95(No.10): p. 1-7.
4. 金子誠一, 証券アナリストのための数学再入門. 2004: ときわ総合サービス.
5. Luenberger, D.G., *Investment Science*. 1998: Oxford University Press.
6. デービッド・G・ルーエンバーガー, et al., *金融工学入門*. 2002, 東京: 日本経済新聞社.
7. Mukherjee, A. and U. Garain, *A review of methods for automatic understanding of natural language mathematical problems*. Artificial Intelligence Review, 2008. 29(2): p. 93-122.
8. Cummins, D.D., *Childrens interpretations of arithmetic word problems*. Cognition and Instruction, 1991. 8(3): p. 261-289.
9. Chang, K.-E., Y.-T. Sung, and S.-F. Lin, *Computer-assisted learning for mathematical problem solving*. Computers & Education, 2006. 46(2): p. 140-151.
10. Bakman, Y., *Robust Understanding of Word Problems with Extraneous Information*. arXiv:math/0701393v1 [math.GM], 2007.
11. Bobrow, D.G., *Natural language input for a computer problem solving system*. Ph.D. Thesis, Department of Mathematics, MIT, Cambridge, 1964
12. Wong, W.-K., et al., *LIM-G: Learner-initiating instruction model based on cognitive knowledge for geometry word problem comprehension*. Computers & Education, 2007.