

学習も容易に進められる。また、もし「確率的言語モデル」なる授業があったなら、教科書として使うのに適しているであろう。

しかし、専門外の読者・初学者にとってはハードな内容になっているかもしれない。そういった人たちは、各章の最初についている入門の節、第7章、解説をまずざっと眺めてみるのも1つの手かもしれない。

さて、自然言語処理に確率的言語モデルを用いることの利点は何か。一番の利点は規則の記述を大量に低コストで行えることであろう。言語の振舞いは非常に複雑なので、この場合はこう処理し、また別の場合はこう処理し、と大量に場合分けされた規則を記述しなければならない。この記述を人手ですべて行うのは容易なことではないし、また人為的ミスも混入することが多い。一方、確率的言語モデルを用いると、場合分けについての指針（たとえば、英語の品詞づけにおける、「ある単語の品詞を決定するには、1つ前の単語と2つ前の単語を見よ」）をある程度与えるだけで、あとは実例（コーパス）から自動的に規則を確率値という形で学習することができる。

一方、現在の確率的言語モデルを用いた自然言語処理の難点を挙げてみる。自然言語処理という分野は言語学などの分野とも密接にかかわってくることであるが、これらの分野の最終目標は何かと問われれば、「人間の言語処理の仕組みを解明すること」となるであろう。この観点からすると、確率的言語モデルはどのような風に位置付けられるであろうか。まずは、以下の日常生活における例をアナロジーとして見てみよう。

ある人が街で観察していて、地面が濡れていれば、傘を差している人が多いと、発見したとする。その観察者は、つまり「地面が濡れれば人は傘を差す」と結論づけるかもしれない。常識的に考えればこの結論はもちろん誤りであり、雨が降っているから、地面が濡れるのであり、傘を差す人が多いのである。

しかし、その結論がまったく無意味かというそうではない。窓もドアも黒塗りで外が見えない商店で働いている店主が、ドアの下から水が入り込んできたのを見つけたとしよう。その店主にとっては「地面が濡れれば人は傘を差す」という規則は十分意味のあるものである。水を見つけた店主が傘をたくさん並べれば飛ぶように売れるであろう。

まとめると、「地面が濡れれば人は傘を差す」という規則は実用的には有意義であるが、常識的に考えてより望ましいのは「雨が降れば人は傘を差す」という、より直接的な規則である。これと同様、確率的言語モデルも、自然言語処理のツールとしては意味があるが、言語処理の仕組みの解明という観点からすれば言語の本質を見落としている可能性があるかもしれない。

したがって、言語の本質の解明のためには、よりよいモデルをどう構築するか、具体的には、何を事象とした確率モデルを構築するか、という点が重要であろう。そのためには、理論言語学などの周辺分野で得られた知見が役に立つであろう。また、それにより精度の上昇という観点からも望ましい結果が得られるかもしれない。このあたりのことは「解説」にも触れているので、そちらも参考にされたい。

ところで、本書は以下の5冊よりなる「言語と計算」シリーズのうちの1冊である。

1. 文法理論の形式化と計算
2. モデル理論の意味論
3. 談話と対話
4. 確率的言語モデル

5. 情報検索と言語処理

古典的な自然言語処理の枠組みでは、自然言語で書かれた文の解析は、統語解析（たとえば、どの単語がどの単語を修飾するかを決定する）・意味解析（たとえば、助詞“by”が「のそばで」「によって」などのどういう意味で使われているかを決定する）・文脈解析（たとえば、代名詞「それ」が何を指しているかを決定する）の順に行われる。これらはそれぞれ本シリーズの1冊目から3冊目に対応している（評者がこれを書いている時点では未刊である）。もちろん、統語・意味・文脈の解析が必ずしも厳密に分けられるものではないが、ここではその話は置いておく（『文法理論の形式化と計算』で触れているであろう）。

一方、本書はこれらを含めて形態素解析・音声認識など言語に関連するあらゆる分野を横断的に、確率という視点から眺めた書である。現状では、確率モデルの研究は形態素解析・統語解析など割合取り組みやすい問題に対して主に行われている。しかし、それらの研究で得られた知見は、意味解析などより難しい問題に対する確率モデルに応用できるであろう。そういった意味で、確率的言語モデルという分野を1冊の本で横断的に眺めるのは価値のあることである。

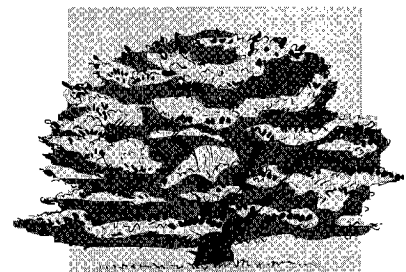
ちなみに、5冊目の『情報検索と言語処理』は、「情報検索」という1アプリケーションの立場から言語処理を眺めた書である。

本書も含めて「言語と計算」シリーズに関する概観的な情報は、

<URL:<http://www.utp.or.jp/shelf/series/gengo.html>>

から得られるので、こちらも参考にされたい。

[光石 豊/東京大学大学院]



会議レポート

EDOC'99

IEEEのComputer SocietyとCommunications Societyなどが共催し、ACMとOMGおよび多数の企業が協賛するEDOC'99 (3rd International Enterprise Distributed Object Computing Conference) が、1999年9月27日から4日間の日程でドイツのマンハイムにて開催された（オーストラリア、アメリカ、ドイツの順で開催され、今年は日本開催となっている）。会議の内容は、<URL:<http://www.wifo.uni-mannheim.de/edoc99/>>で参照できる。

本国際会議の名称が第2回まではワークショップであったように、参加者の討論と交流を重視した会議である。たとえば各論文発表セッションでは、発表者をパネリストとする20分程度のミニパネルを行う。また参加者も150名以上は多くしないなどの方針で運営している。会議は分散オブジェクト技術を中心テーマとするが、“Enterprise”のキーワード

で他のオブジェクト指向関連の国際会議と一線を画している。すなわち、企業向けシステムに必要なと思われる技術やアプリケーションをはじめ、大規模性や信頼性等を議論の対象とする会議である。今年の論文の採択は28件で約50%の採択率であった（レフリーは各論文に3名）。今回は特にアメリカや日本からの論文投稿が少なかったのが残念である。プログラムの構成は、招待講演2件、チュートリアル6件、論文28件、パネル2件で、参加者は100名と前年より少なかった（前回は125名）。

招待講演は以下の2件である。

- “Patterns for Concurrent and Distributed Objects”, Frank Buschmann (Siemens Corporate Technology)

Buschmannはパターン研究で著名である。講演では、分散オブジェクトシステムにおけるパターン技術の開発の重要性と具体例を説明した。現在、多くのパターンの研究がデータベースアプリケーションなどの対話的システムを対象としているが、交通システム、プラントシステム、医療システムなどの分散システムには多くの技術課題があり、パターンによりそれらの解決法を整理することが重要であることを強調した。

- “Unleashing the Power of Distributed Enterprise Information Systems”, Trygve M.H. Reenskaug (University of Oslo)

講演者が重要と思っているソフトウェア開発やオブジェクト指向に対する考え方を説明。オブジェクト指向システムは、クラスでなくオブジェクトで考えるべきである。最近の技術やシステムは複雑化する傾向があるが、もっと単純で重要なアーキテクチャやコンセプトを定義するよう努力すべきである、と主張した。

チュートリアルは以下の6件である。

- (1) “Enterprise Modeling Using UML, MOF, and XML”, S. Iyengar (Unisys)
- (2) “What's New in CORBA 3.0”, Jon Siegel (OMG)
- (3) “Catalysis: Modeling for Component Architectures”, Alan Wills (Tireme)
- (4) “Using UML for Architecture Description”, Derek Coleman (HP Product Generation Solutions/Consulting)
- (5) “An Introduction to the Enterprise JavaBeans”, Andreas Liebscher (MGM)
- (6) “Enterprise CORBA”, Dirk Slama (IONA)

チュートリアル (1) は、今後の分散システムの開発環境と運用環境においてリポジトリが果たす役割と重要性について解説した。講師はOMGで本分野のChairとして活躍中。リポジトリには多くの定義があるが、データベースに関する意味情報（メタデータ）を保持する拡張可能なデータベースアプリケーションであると説明。リポジトリは情報システムの拡張や統合などにきわめて重要な役割を果たす。今後、データベースハウスも迅速な更新と基幹アプリケーションとしての要求が増すことにより、リポジトリ技術が重要となると説明した。チュートリアル (4) は、ヒューレットパッカード社のオブジェクト指向開発技法Fusionで著名なColemanが講師である。ソフトウェアアーキテクチャの開発の重要性と、

UMLを使った開発プロセスと記法上のポイントを説明した。多くのドキュメントを早い時期から完結させるのではなく、重要な部分を見つけてそこにフォーカスし、適正かつ必要最低限のドキュメントを揃える必要性を強調した。

論文発表は28件で、10のセッションに分けて発表された。

- ビジネスオブジェクトの用語と技術の整理、E-Commerce、製造業におけるプロセスと製造物を統合するメタモデルに関する発表が2件
- マルチメディアアプリケーションにおけるJavaリフレクション応用に関する発表が2件
- システムの要求をISO標準のRM-ODP (Reference Model for Open Distributed Processing) の概念やUMLを使って表現するもの4件
- ODPの概念をベースとした仕様記述に関するもの2件
- グループウェアやワークフローの提案や実装に関するもの2件
- さまざまなアプリケーション向けのフレームワークやミドルウェア技術に関するもの4件
- 仮想企業に関するもの3件
- アーキテクチャに関するもの3件
- ミドルウェアのサービスレベルの提案に関するもの4件
- CORBAの最近の改善点に関する評価やコンポーネントに関するもの2件

である。また、パネルは以下の2件である。

- “Methodologies for the Development of EDOC Systems: Is modelling worth the effort?”

モデリングの重要性や何を指針として作業すべきかが、各パネリストの研究の背景をベースに討論された。Colemanは技法が提示する手順にただ忠実に従うのではなく、リスク主導、ビジネス主導、そしてアーキテクチャ主導で開発を進めるべきであり、インターネット時間（超短期間）での製品企画・開発が求められる現在、狭い範囲で詳細にモデル化するのでなく、広い視野で重要点の見落としがないようにモデル化し開発することが大事である、と述べた。

- “Imposing access control and business rules over an enterprise: How important is it? How can it be done?”

このパネルでは、ビジネスルールの定義や見方（システム側かビジネス側か）があいまいで議論がかみ合わない場面が見られた。今後の会議でもこの議論は続きそうである。

会議はマンハイム城の敷地内にある大学で行われたが、3日目の夜は、会場から20kmほど離れたハイデルベルグ城の中でカンファレンスディナーが開催され、ライトアップされた城と庭を眺めながら、おいしいワインを飲み、楽しいひとときを過ごした。なお次回は情報処理学会が主催し、ソフトウェア工学研究会の運営で9月に幕張で開催される<URL:http://edoc.ae.keio.ac.jp/edoc/>。ぜひ、論文投稿あるいは会議に参加して多くの研究者・技術者と交流していただきたい。

(佃 軍治/日立製作所システム開発研究所、
上原三八/富士通研究所)