論文の内容の要旨

氏名：野　本　秀　樹
博士の専攻分野の名称：博士（工学）
論文題名：機能共鳴分析手法（FRAM）を用いた不特定多数の人工知能エージェントによる自由行動の安全化に関する研究

人工知能を搭載した不特定多数のエージェント（各自の意思決定原理機構に基づき動作する自律的な主体）が，巨大な IoT（Internet of Things）システムを構成しつつ自由に行動する現実・仮想の空間をいかに安全なものにし得るか，という課題は，特に自動運転等の分野で早急に解決されなければならない問題となってきている。例えば，一般道路で自動運転を行う場合，道路上には，様々な人・他車・モノが存在し，時に予測困難な挙動を示すことがある。人間にとっては難なく対処できる状況であっても，ソフトウエア制御による自動運転システムを設計する上では，難易度が極めて高いことが多い。常に想定外の状況が発生する可能性があるからである。そうした問題に対しては，従来のルールベースの制御ロジックではなく，人工知能を使用するという考え方を自動車メーカー自らが表明し始めている。一方，ソフトウエアにより高度に自動化されたシステムの安全化については，従来以下の戦略がもっぱら採用され，成果を上げてきた。

・想定されるハザードに対して，ハザードを発生させる原因を識別する
・ハザード発生原因ごとに発生を防止するための安全制約を適用する
・ハザード発生原因ごとに発生を検知し安全化するための制御を作りこむ

これらの方法は，制約，すなわちバリアを設けること，あるいは危険状態を検知して積極的に制御をかけることが有効であるという前提条件の下で成り立つ。

しかし，多数の自然知能（人）が自由に行動する現実空間は，必ずしも上記の方法で安全化されてきたわけではない。例えば，本論で分析する東京駅のコンコースにおいては，人の行き来を制約するルールや，危険を検知して人の流れを制御するような機構は設けられておらず，一見するところ，単なる通路でしかない。つまり，東京駅の安全性は，既存の安全制御のスキームとは異なるものによって達成されている可能性が高い。

人工知能の問題を考察するにあたり，既存の安全制御の方法に縛られることなく，自然界に既に存在している例から将来の安全化策のヒントを得ることは，有効であると考えられる。

本研究では，東京駅のコンコースがどのように安全に保たれているのかを分析し，次世代の人工知能安全につながる指針を得ることを目的に，機能共鳴分析手法 FRAM（Functional Resonance Analysis Method）を用いて，システムの成功要因（システムが安全に目的を達成できる要因）・リスク要因を識別するとともに，同手法によって生成されたモデルの妥当性を検証するためのシミュレーションを行う。

本研究の概要は以下のとおりである。

第 1 章「序論」では，研究の背景を明らかにし，研究の概要を述べた。

第 2 章「従来の安全解析手法の問題点と新しい安全解析手法」では，現在主流となっている安全解析手法を，本研究の対象であるインテリジェントな不特定多数のエージェントによる自由行動に適用する際の問題点について述べた。

第 3 章「FRAM 分析」では，本研究において使用する FRAM の概要及びこれを使用したモデリング方法，研究対象のモデリング結果，モデルの評価方法，研究対象のモデル評価結果について述べ，以下の知見を得た。

東京駅における各歩行者の安全は，交通整理や信号制御などのトップダウン制御ではなく，歩行者各自が別々に実施しているボトムアップな歩行戦術の絶え間ない更新によって実現されている。また，

東京駅という場が提供している安全上の機能は，各歩行者に対して目的地をあらゆる場所で確認できるよう,おびただしい量の行き先案内板によって表示するという点にある。目的地の安定的な提供と，歩行のための戦術の絶え間ないボトムアップな更新が，システム全体の安全を築き上げている。このようなモデルから分析された特徴から，以下のように成功要因とリスク要因が識別された。

(1) 成功要因

一旦ゴールが設定されると，ボトムアップに練り上げられる戦術にしたがって皆が歩き続けることができる。戦術は,常に自動的に参加者全員によって環境変化に適応して更新し続けられる。戦術を決めるのは，ルールや交通整理によるトップダウンプロセスではなく，参加者によるボトムアッププロセスである。

ゴールは環境に影響されず，常に安定的に提供されている。

(2) リスク要因

ゴールの提供が滞ると，歩行が停止するため，新規流入者を受け入れることが困難となり，パニックに至る。

第 4 章「シミュレーション」では，FRAM モデルからシミュレータを作成する方法及びシミュレーション実行の分析結果を述べた。

シミュレーション実行結果に対する分析の結果，東京駅における歩行の特徴として，以下が識別された。

(1) 自由に移動可能な自分の周りのスペースに，一人分の空きスペースがある場合，そこに移動可能とする閾値（Minimum Distance = ″1″）の条件で実行されたゴール到達時間や停止回数が，現実の東京駅の状況に近い。

(2) Minimum Distance = ″1″のケースにおけるような安定した系の性質が見られる領域では，系の安全性（衝突・停止回数）と経済性（ゴール到達時間）は強い相関を有しており，安全であればあるほど経済的になるという好循環が，東京駅コンコースの特徴となっている。

(3) 安全性，経済性が共に良好な状態を維持できている安定状態から，構内の多くの歩行者が停止する麻痺状態に陥る状態へ遷移する条件は，Minimum Distance と歩行者数の組み合わせで決定される。

(4) 安全であればあるほど経済性が高まるという東京駅のコンコースの持つ性質は，従来の制御系システムの安全設計では実現が困難な長所であり，「本質安全」と呼べる優れた性質を備えている。

(5) 東京駅のコンコースには自己組織化による自律的な安全化メカニズムが働いていると考えられる。東京駅の歩行者は，それぞれの目的にしたがい利己的に行動しているため，系として統一行動はとっておらず，系のエントロピーは非常に高い乱雑なシステムである。しかしながら，コンコース内で発生した異常に対して，あたかも優れたリーダーに統率されているかのように対応できるため，大きな事故が発生することはまれである。それは，バイアレク等により見出された鳥の群れに見られる相転移現象が，群れとして高いエントロピーを有した状態から発生するという説と通じるものである。

第 5 章「対象システムの安全化に関する考察」では，本研究に強く関連する次世代安全理論について述べ，今後の発展の可能性について言及した。

第 6 章「結論」では，本研究で得られた主要な成果をまとめ，本研究の意義と課題について整理した。

本研究は，今後急速な技術進歩が予測されるロボットや自動車の自動運転といった自律したエージェントが，安全性を損なうことなく共存できるための分析手法の確立を意図したものであった。安全を積極的に制御するプレイヤーがいないマルチエージェントシステムが，どのように安全化されているのかを FRAM により可視化するという試みは初めてのものであり，シミュレーションを用いてのケーススタディでは,「自己組織化による先天的安全性の創発が最大の特徴」との興味ある結果が導き出された。

本研究においては，自由度の高い FRAM モデルを，トップダウン型安全制御モデルである STAMP モデルに写像するという独創的な方法を考案した。ホルナゲルの提案による FRAM は，システムの表現（成

功要因）はできるが，安全解析の方法論は提示されておらず，どのように利用するかが課題であるとして限界が指摘されていた。しかし，本研究で考案した方法論により，FRAM が単なる表現手段としてではなく，安全性分析のツールとしても利用できることが実証された。

今後，多数の自動運転車両が人間や非自動運転車と混在して自律走行する社会や，多くの支援ロボットが人間と共存して動き回る社会の到来が予見されるが，本研究がそのような社会における安全のあり方の検討に貢献できるよう，研究の深度化に努めていきたいと考えている。