東海大学紀要情報通信学部
Vol.10,No.1,2017,pp.114-116

トピックス

安全性についての課題ー価値判断とAIの使用について

小林　洋*1

Issues of Safety—Estimate of Safety and Decision Using AI

by

Hiromi KOBAYASHI*1

(received on Mar.24, 2017　&　accepted on Jul.13, 2017)

あらまし

現代社会においては，新しい技術が出現する度にその技術により新たな安全性の課題が生み出され，リスクに応じた安全対策が必要となる．本稿では，まず安全性という用語の意味を，基本となる国際規格を基に解説する．次に，安全対策を立てる上で検討する必要があるリスクの評価について述べ，そこで現れる価値判断の問題を示す．最後に，最近注目されている自動車の自動運転での人工知能（AI）の使用について，安全性の観点から課題について述べる．

Abstract

A new technology produces new issues of safety. First, this paper presents the meaning of safety based on international standards. Next, we address the problem of risk assessment in order to take safety measures. Finally, we show the decision-making problem of automated vehicle using AI.

キーワード:安全性，許容可能，リスク評価，自動運転，人工知能
Keywords: *safety, tolerable, risk assessment, automated vehicle, AI*

1. はじめに

現代社会においては，新しい技術が出現する度にその技術により新たな安全性の課題が生み出される．新しい製品やシステムについては，開発途中ではあまり明確に認識できなかった，或いは誰もが想定しなかったような事故が起きると，安全上の問題が議論されるようになる．安全性というのは，人間にとっては本能的とも言える欲求で，誰でも良く解っている事のように思えるのだが，議論になると話がかみ合わない事がある．場合によっては，安全性という言葉の意味から認識が異なる事がある．そこで本稿では，まず議論する上で前提となる安全性という用語の現代的な意味を，基本となる国際規格を基に解説する．次に，安全対策を立てる上で検討する必要があるリスクの評価について述べ，そこで現れる価値判断の問題を示す．価値判断は人により異なる場合があるため，何に対してどの程度の安全対策を立てるのかの合意の形成が必要となる．最後に，最近注目されている自動車の自動運転での人工知能（AI）の使用について触れ，安全性の観点から課題について述べる．

2. 安全性における許容可能という考え方

日常的に安全といっている言葉は，国際規格においては英語の safety と security の二つの用語が対応し，国内規格などでは前者を安全，後者をセキュリティと呼んで区別する．概念としては，前者は正しく運用して損傷させる意図などないにも関わらず機材の故障や操作ミスのためにシステムにとって悪い振る舞いが発生し損傷が生じる事，例えば機材の故障やオペレータの操作ミスによる事故の防止などを対象とする．一方，後者は意図的に損傷させようとする場合に生じる事，例えば最近深刻な問題となっているインターネットでのウィルス等のマルウェアによる被害の防止等を対象とする．但し，損傷を防ぐという共通の目的の対策を考える上では，両者について総合的に考える必要がある．安全やセキュリティについては，WTO（World Trade Organization）の TBT（Technical Barriers to Trade）協定に基づき，製品やシステムの輸出入における技術的障壁を取り除く必要性から，国際規格とそれに対応する国内規格が定められている[1-2]．安全性については，まずガイドラインとしての国際規格 ISO/IEC Guide51 があり，それに対応する日本の規格 JIS Z 8051 では，安全についての定義は「許容できないリスクがないこと．」とされ，信頼性用語の国際規格 IEC60050(191)に対する JIS Z 8115 では，安全は「人への危害又は資（機）材の損傷の危険性が，許容可能な水準に抑えられている状態」とされている．なお，国際規格では資（機）材は asset，許容可能は tolerable(規格によっては,acceptable)と書かれている．一方，セキュリティについては，ISO/IEC 15408 に対応する JIS X 5070(或いは,Common Criteria（セキュリティ評価基準）の日本語版）に記されている．但し，こちらでは asset の部分は資産と訳されており，資産とは何者かによって価値が認められているもので，価値とは非常に主観的なものであるとの旨が述べられている．安全性の定義で注目すべき

*1 情報通信学部情報メディア学科 教授
School of Information and Telecommunication Engineering, Department of Information Media Technology, Professor

点は，安全という状態が損傷の危険性が許容可能な水準となっていて，危険性がゼロとはなっていない事である．製品やシステムの開発において，危険性をゼロとすることは，コスト的にも時間的にも現実的に不可能であるため，許容水準以下に抑えられていれば良しとする考え方である．安全性については，共通のものと分野ごとの様々な国際規格と国内規格が存在するがその概説的なものは，例えば文献 1-2)などに記載されている．なお，最近では，安全に加え，安心という心理的な要素も加え，安全と言われている事を信頼したいという意味であろうか，安全・安心と言う言葉がしばしば使われているが，安心という心理面の評価に対しての工学的な規格は存在しない．

3. リスクの評価における価値判断の問題

安全性にせよセキュリティにせよ，安全対策については，一般的には，危害の発生確率と危害の程度の組み合わせで表されるリスク（risk）の評価に対応して行われる[1-2]．危害の程度は，しばしば量的に比較しやすいように金額により損害額として表し，リスク＝危害発生確率×損害額とされる傾向があるのだが，危害の程度を果たして損害額を掛けただけとして良いかは議論の余地がある．ところで，危害を評価するには，まずは危害に対する価値判断をしなければならない[3-5]．価値というのは，組織での立場によって異なる場合がある他，主観的で人によって異なる場合がある．特に，複数の人や資産が対象になる場合，それらについての価値判断をしなければならないことが，安全対策を考える上で難しい点である．対象とするシステムが単純な単一の評価尺度で表される階層構造のみから構成されているのであれば，システムの上位の方が下位よりも優先することで良いのだが，システムが単純な階層構造で無く構成要素のお互いの関係が明らかで無いような場合には，価値判断に関する合意の形成が必要になる．このような研究は，いくつかの分野で研究されているようだが，情報系の分野では，分散人工知能やマルチエージェントシステムと呼ばれる分野等で研究が行われている[6-7]．一般的に合意の方法としては，先ず，予めルールや優先順位を決めておくという方法がある．ルールや優先順位を予め決めておくことにより，緊急時の場合に速やかに処理を行うことができる．別の方法として，利害関係者同士の交渉に委ねるという方法がある．但し，多数の利害関係者間の交渉は一般的には時間がかかる．その他に，第三者に委ねるという方法があり，これには第三者による強制的な指令または調停に委ねる等の方法がある．　なお，リスク評価における危害発生確率の方も問題があり，従来型の交通事故のように統計データがある場合には，妥当と思われる確率が出せるが，新規の技術の場合のように経験的な統計データが乏しい場合には妥当と思われる確率を出すのが難しくなる．

4. AIの判断と安全性

最近の安全性についての話題の一つに自動運転技術に関することがある．2013年5月には，将来を見据えて，米国道路交通安全局（NHTSA）から自動運転に関する提言（preliminary statement of policy concerning automated vehicles）[8]がなされており，そこには自動化に関して5段階のレベルが示されている．現在，自動運転については世界中の主要な自動車メーカやグーグルなどが競って開発を進めており，2020年ごろを目途に実用化を目指している．しかし，2016年5月に米国テスラ・モーターズ社製の電気自動車が，オートパイロットモードでの運転中に，高速道路で大型トレーラと衝突しドライバーが死亡するという事故が発生したことは，ブームへの注意喚起となった[9]．また，近年人工知能（AI）が第三次のブームとなっており，特に，深層学習（ディープラーニング）技術[10-12]が，予測や分類を行うのにブレイクスルーとなる技術として脚光を浴びており，そう遠くない将来において自動運転などにも取り入れられるのではないかとの期待とも怖れともとれる論調が雑誌記事[13]などには見られる．但し，テスラの自動車の場合は限定した機能のみの自動化であり，完全自動運転ではないので，それからはやや飛躍するのであるが，自動運転での事故の際の判断に関わる問題として，2016年10月にいわゆるトロッコ問題と言われる「直進すれば子供たち（通行人）が死に，回避すれば運転手が死ぬ．」というような状況での判断についてのドイツのメルセデス・ベンツ社のある担当者の回答が話題となった．それは，完全自動運転では運転手を助けることを優先するという回答[14]であるが，自動車メーカの担当者としては当然の回答のように思われる．自動運転のレベルの中で最も高レベルに当たる完全自動運転でディープラーニングを用いるようになるかどうかは解らないが，ある種のAIと呼んでも良いような高度な処理を行う情報システムでの判断とそれに基づく制御を行うことになるのであろう．この場合のAIの判断には，事前のメーカの人間によるAIの訓練時の価値判断が含まれている事になるはずである．しかし，ディープラーニングの場合には，人間の思いもよらない解を出すことがあり，囲碁の対局の場合には，トッププロ棋士でもその時は理解できなかったが，後でそれが優れた手であることが判明するという事態が発生した[15]．囲碁のようなゲームの場合にはそれでも良いのだが，自動運転での危険の回避において人間の思いもよらぬ解を出された場合に，判断をそれに委ねて良いものかは疑問である．また，ディープラーニングにおいては，今のところ，なぜそのような解を出すに至ったのか途中経過を追跡するのが困難であるため，この技術を自動運転に取り入れようとすると，自動車関連の機能安全規格 ISO26262 においては，セーフティケース（Safety Case）[16]と呼ばれるシステムが安全に作られていることの証拠文書の提出が義務付けられていることから，そこに処理過程がブラッ

クボックス化してしまうディープラーニングが根拠として使われることは，今のところはできないではないかと思われる．AI による判断の問題について，自動運転において発達している航空機の分野で考えてみると，航空機と自動車では機能や性能の他，環境や運転者（操縦者）の訓練の程度が違うために，比較は困難なのかもしれないが，参考までに二つのケースを取り上げることにする．まず，航空機の着陸時の問題として，人間とコンピュータ（今風に言うなら AI）の判断が異なる場合どちらを優先すべきか，という問題が以前から問われて来た．これは，悪天候の空港などの着陸の場合などでは人間は錯覚したり，また緊急の際には判断を誤る事があるのでコンピュータに任せた方が良いという考え方と，臨機応変に対処するために熟練パイロットの判断に任せた方が良いとする考え方の問題で，俗にエアバスはコンピュータ優先なのに対して，ボーイングは人間優先と言われて来た[17]．但し，エアバスにおいても 1994 年の名古屋の中華航空 140 便墜落事故の教訓などから，自動を優先としながらも自動操縦から手動操縦への切り替えが容易に行えるように改修されている[18]．次に，航空管制の問題として，空中衝突防止装置（TCAS）と航空交通管制官（ATC）の指示が違う時には，どのように解決するかという問題があるのだが，これは，今では常に TCAS 優先とされている[19]．これは，２機の航空機の TCAS Ⅱ同士においては回避指示の協調が自動的に行われ，一方が降下なら他方は上昇と指示が出されるためで，ATC がこれに矛盾する指示を出すことによる事故を防ぐためである．つまり，この二例で見ると個々のケースによって対応が異なっており，自動車の場合にも，個々のケースごとの対応を積み重ねて行くしかないように思われる．

5. おわりに

本稿では，最初に安全性という用語の意味を，国際規格を基に解説した．安全性の基本的な考え方は，危険性ゼロという絶対安全を基にすると人間活動に支障を来すことになるため，損傷の危険性の許容可能水準を基に安全性を評価するという事である．次に，安全対策を考える上で基となるリスクの評価について述べた．リスクは危害の発生確率と危害の程度の組み合わせと定義されているが，危害の程度の評価には価値判断が伴い，価値判断は人によって異なるため，何に対してどの程度の安全対策を立てるのかの合意の形成が必要となる事が課題として上げられる．最後に，最近開発が進んでいる自動車の自動運転を取り上げ，そこでの人工知能（AI）の使用について，安全性の観点から課題について述べた．最近話題のディープラーニングについては，今のところ内部処理の追跡が難しくブラックボックスとなる事が安全性の点からは問題である．AI と全て呼んで良いのかどうかは解らないが，高度な処理を行う情報システムの判断と人間の判断との優先問題については，個別のケースによって対応が異なるように思われる．一般論としてコンピュータと人間の判断のどちらかを優先するかという事のみを議論する事は，事実に基づかない空論に陥ってしまう可能性があり，自動車の場合にも，個々のケースごとの対応を積み重ねて行くしかないように思われる．

参考文献

1)向殿政男監修，安全の国際規格１：安全設計の基本概念，２：機械安全，３：制御システムの安全，日本規格協会，2007.

2)佐藤吉信，機能安全の基礎，日本規格協会，2014.

3)小林洋，ソフトウェア開発で対象とする安全性について，日本信頼性学会第11回研究発表会発表報文集，日本信頼性学会誌，pp.325-326，2003.

4)H.Kobayashi,Safety Concern in System Development, Knowledge-Based Software Engineering, Vol.108, IOS Press, pp.311-318, 2004.

5)小林洋，システムの安全性とは何に対する安全性なのか？，ソフトウェアシンポジウム2012(SS2012)，ソフトウェア技術者協会，pp.3.1-3.7，2012.

6)G.Weiss編，Multiagent Systems 2nd edition, MIT Press, 2013.

7)S.Russel, P.Norvig, Artificial Intelligence: A Modern Approach (3rd Edition),Pearson, 2009.

8)米国道路交通安全局（NHTSA），自動運転に関する提言（preliminary statement of policy concerning automated vehicles），https://www.nhtsa.gov/, 2013.

9)日本経済新聞電子版，http://www.nikkei.com/, 2016/7/1.

10)松尾豊，人工知能は人間を超えるか，KADOKAWA，2015.

11)神蔦敏弘編（人工知能学会監修），深層学習，近代科学社，2015.

12)岡谷貴之，深層学習，講談社，2015.

13)小林雅一，自動運転車の事故は「原理的に」避けられない!? AI 技術の死角 ，http://gendai.ismedia.jp/articles/-/49777, 2016/9/22.

14)AutoExpress:http://www.autoexpress.co.uk/mercedes/97345/mercedes-autonomous-cars-will-protect-occupants-before-pedestrians

15)大橋拓文，伊藤毅志，“Master”の衝撃，pp.160-165，人工知能学会誌，Vol.32，No.2，2017.

16)所眞理雄編，DEOS:変化しつづけるシステムのためのディペンダビリティ工学,p.25，近代科学社，2014.

17)福田収一，大きく変る信頼性，電子情報通信学会誌,pp.1022-1025，Vol.89，No.12,2006.

18)中華航空公司所属 エアバスインダストリーA300B4-622R 型 B1816 愛知県名古屋空港,航空機事故報告書，96-5-B1816 運輸安全委員会，1996.

19)TCAS Version 7.1 Eurocontrol, http://www.eurocontrol.int/ articles/tcas-ii-version-71