Информационная безопасность(ИБ) экспертных систем и систем искусственного интеллекта

Баранов Александр Павлович, Д.Ф.-М.Н., академик Академии Криптографии Российской Федерации, АО «Аналитический Центр»,

Баранов Юрий Александрович, К.Т.Н.

Экспертные системы(ЭС) и Системы искусственного интеллекта (СИИ).

1. ЭС- обработка данных и информации, адаптивные системы.

СИИ-генерация новых знаний, на основе сформированной естественным интеллектом базы знаний и методов получения новых

2. Принципиальная узость специализации ЭС и СИИ в областях применения, для которых их разрабатывали. Китайский взгляд

3.Опасные направления нападения:

а) Специальные представления данных и их структуризации;

б) Представление и декомпозиция(разделение на отдельные фрагменты) знаний и их использование для обработки данных;

в)Принципы вывода новых знаний и использования данных;

г) Группы экспертов по формированию алгоритмов обработки данных, систем адаптации, базы знаний, по осмыслению получаемых результатов

Направления применения ЭС и СИИ в критических технологиях

1. ЭС в управлении технологическими процессами производств. ЭС реального времени. Обработка и осмысление данных
2. ЭС постановки диагноза и предложение метода лечения. СИИ исследований биологических, физических, химических процессов
3. ЭС оценки финансовых и экономических ситуаций. СИИ прогнозирования на основе обработки предыдущих данных
4. ЭС в составе робототехнических комплексов реального времени и реализации реакции. Военные применения
5. СИИ для поиска новых знаний, науки, процедур самообучения и адаптации

Угрозы при применении ЭС и СИИ

1. «Подсовывание» ложных, тенденциозных данных, искажение выдаваемых, нарушение конфиденциальности(К) деятельности
2. Включение в состав групп экспертов «неправильных» специалистов, имеющих возможность повлиять на сформирование недостоверного представления знаний или искажение алгоритмов расчетов
3. Использование расчетных процедур или их нейронных сетей не по предназначению
4. Мифологизация возможностей ЭС и СИИ для продвижения научно отравленных приманок и отвлечения на негодный объект исследований

ИБ собственно самих ЭС и СИИ

1. Обеспечение К, Ц, Д программного обеспечения(ПО) и аппаратных средств, составляющих ЭС или СИИ

2. Проблема доверенной аппаратной платформы в условиях отсутствия процессорной базы менее 180нм

3. Обеспечение безопасного режима разработки ЭС и особенно СИИ. Проектирование и разработка собственного ПО представления предметной области , базы знаний, получения нового знания, обучения экспертов передаче знаний и т.д

4. Защита ЭС, находящихся в «облаке» при интенсивном многопользовательском массовом режиме доступа. Пример: поиск гражданами маршрутов в геоинформационной системе

Направления ИБ, используемых,входящих и выдаваемых данных

1.Традиционная защита данных К,Ц,Д при обработке, хранении выдачи результатов

2. Обеспечение ввода данных различных типов (аудио, видио, графических, текстовых) с адекватной точностью

3. Верификация данных при обучении и разработке ЭС или СИИ, проверка соответствия данных модели представления предмета исследования или взаимодействия

4. Структуризация, рубрикация и разметка «сырых» данных при использовании систем хранения типа «озера данных»

5. Обработка данных в искусственных нейронных сетях по верифицированным ситуациям и моделям действительности

ИБ базы знаний при ее формировании и эксплуатации

1. Представление знаний(ПЗ) в выделенной четко очерченной области. Формирование группы доверенных экспертов- специалистов по ПЗ
2. Разработка эффективной системы обучения специалистов области способам представления своих знаний для включения их в базу. Контроль противоречивости суждений экспертов
3. Обеспечение адекватности экспертов на отсутствие ангажированности и предвзятости
4. Разработка и оценка эффективности методов получения нового знания. Формирование позитивного климата рассмотрения и оценки экспертами полученной новизны.
5. Обеспечение сохранения конфиденциальности и регистрации приоритетности за системой новых результатов. Недопущение присвоения нового знания экспертами-оценщиками

Выводы

1. Предлагается принципиально различать два типа интеллектуальных систем: обработки данных- ЭС и получение новых нетривиальных знаний- СИИ
2. ИБ ЭС и СИИ состоит из двух составляющих: ИБ алгоритмов, ПО, аппаратного обеспечения и ИБ процессов разработки и эксплуатации
3. Эффективность и надежность ЭС и СИИ в значительной степени обеспечивается специальной работой с экспертами, отбраковкой и проверкой их релевантности
4. Применение не верифицированных ЭС и СИИ в непредусмотренных разработчиками задачах может приводить к негативным последствиям и неправильным решениям, в следствии неадекватного моделирования окружающей среды
5. Обеспечение ИБ при разработке и эксплуатации компьютерных интеллектуальных систем является обязательной, специфической составляющей этих процессов