ВСТУП

Автоматизована ідентифікація багатоатомних молекул за допомогою методів машинного розпізнавання інфрачервоних (ІЧ) спектрів є важливим напрямком досліджень і використання в різних галузях, таких як хімія, фармація, біологія та матеріалознавство. Застосування методів машинного розпізнавання дозволяє ефективно обробляти великі обсяги даних ІЧ спектрів, покращуючи швидкість і точність ідентифікації молекул. Автоматизована обробка даних також спрощує інтеграцію цих методів у виробничі та лабораторні процеси, роблячи їх більш доступними та ефективними.

Метод ІЧ спектроскопії дозволяє отримувати характеристичні вібраційні спектри для різних хімічних з'єднань. Автоматизована обробка цих спектрів за допомогою алгоритмів машинного навчання дозволяє точно ідентифікувати хімічні сполуки у зразках. Використовується для аналізу харчових продуктів, визначення їхнього хімічного складу та виявлення можливих забруднень. Автоматизована ідентифікація дозволяє забезпечити високу точність та швидкість аналізу продуктів. У фармацевтичній промисловості важливо контролювати якість вироблених лікарських препаратів. Автоматизована ідентифікація молекул за ІЧ спектрами дозволяє швидко та ефективно визначати наявність необхідних компонентів та контролювати їхні концентрації. В ІЧ області можна вивчати вібрації біомолекул, таких як білки, ДНК і РНК. Це важливо для дослідження структури та функцій біомолекул, що є важливим у біології та медицині. Автоматизована ідентифікація за ІЧ спектрами сприяє удосконаленню процесів синтезу та контролю якості матеріалів.

1. ОСНОВИ ІНФРАЧЕРВОНОЇ СПЕКРОСКОПІЇ

1.1 Призначення та принцип роботи інфрачервоної спектроскопії

Інфрачервона спектроскопія(ІЧ спектроскопія) — різновид молекулярної оптичної спектроскопії, оснований на взаємодії речовини з електромагнітним випромінюванням в ІЧ діапазоні: між червоним краєм видимого спектра (хвильове число 14000 см-1) і початком короткохвильового радіодіапазону (20 см-1).

ІЧ спектри виникають при поглинанні ІЧ випромінення на частотах, що збігаються з деякими власними коливальними і обертальними частотами молекул або з частотами коливань кристалічної ґратки. ІЧ спектри отримують за допомогою спектрометрів різних типів, робочий діапазон яких знаходиться в межах так званої фундаментальної ІЧ області (400 см-1 — 4000 см-1).

Інфрачервона спектроскопія належить до абсорбційних спектроскопічних методів. Базується на здатності молекул поглинати ІЧ-випромінювання із збільшенням коливальної і обертальної енергій ковалентного зв’язку.

ІЧ-ділянка спектра знаходиться приблизно від 0,8 до 200 мкм (довжини хвиль) або від 12 000 до 50,00 см-1 (хвильові числа). Цю ділянку можна розбити на декілька більш дрібних інтервалів відповідно до різних типів коливань, що мають місце, і відповідно до апаратури для їх спостереження.

Довжинам хвиль від 0,7 до 2,5 мкм відповідають валентні коливання між атомами Гідрогену та іншими атомами. Фундаментальні коливання відмічають між 2,5 і 50 мкм. Спектри поглинання у цій ділянці характерні для окремих часток, особливо інтенсивно використовують інтервал 2,5–15 мкм, де з’являється велика кількість піків поглинання функціональних груп. Смуги поглинання в інтервалі 15–50 мкм відповідають деформаційним коливанням важких атомів і груп атомів. Поглинання у далекій ІЧ-ділянці відповідає НЧ-коливанням та обертанням і аналітично широко не застосовуються.

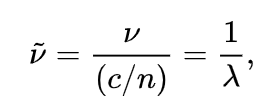
ІЧ-спектри представляють двома способами: по осі ординат відкладають пропускання у відсотках, а вісь абсцис може бути калібрована у мкм (довжини хвиль) або см-1 (хвильові числа) залежно від вибору апаратури для реєстрації спектрів. ІЧ-спектри служать джерелом інформації про структуру молекулярних сполук і широко застосовуються для ідентифікації органічних речовин.

Інфрачервона спектроскопія із хвильовими числами менше 100 см-1 при поглинанні органічною молекулою перетворюється на енергію обертання. Поглинання є квантованим, тому обертальний спектр молекул складається із дискретних ліній.

Випромінювання в інтервалі 10 000-100 см-1 при поглинанні перетворюється органічною молекулою на енергію коливань. Це поглинання також квантоване, але коливальний спектр складається не з ліній, а зі смуг, оскільки кожний коливальний перехід супроводжується змінами обертальних станів.

Поглинання смуг у спектрах позначається через хвильові числа �¯, які вимірюють у зворотних сантиметрах см-1; ця одиниця є прямо пропорційною енергії коливання, а шкали спектрометрів є лінійними відносно см-1. Хвильові числа також називають "частотами". Це неправильно, оскільки хвильові числа �¯ у см-1 дорівнюють 1⋅104/� у мкм, а частоти � у Гц дорівнюють �/� у см, де � - *c* швидкість світла.

ІЧ-спектроскопія базується на явищі поглинання хімічними речовинами інфрачервоного випромінювання з одночасним збудженням коливань молекул. Інфрачервоне випромінювання є [електромагнітною хвилею](https://uk.wikipedia.org/wiki/%D0%95%D0%BB%D0%B5%D0%BA%D1%82%D1%80%D0%BE%D0%BC%D0%B0%D0%B3%D0%BD%D1%96%D1%82%D0%BD%D0%B5_%D0%B2%D0%B8%D0%BF%D1%80%D0%BE%D0%BC%D1%96%D0%BD%D1%8E%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F) та характеризується [довжиною хвилі](https://uk.wikipedia.org/wiki/%D0%94%D0%BE%D0%B2%D0%B6%D0%B8%D0%BD%D0%B0_%D1%85%D0%B2%D0%B8%D0%BB%D1%96) λ, [частотою](https://uk.wikipedia.org/wiki/%D0%A7%D0%B0%D1%81%D1%82%D0%BE%D1%82%D0%B0) ν і хвильовим числом �, які пов'язані такою залежністю:



де c — швидкість світла, а *n* — показник заломлення середовища.

У спектроскопії поглинання, частковим випадком якої є ІЧ-спектроскопія, відбувається поглинання молекулами фотонів певної енергії, яка пов'язана з частотою електромагнітної хвилі через сталу Планка:



При поглинанні фотона відбувається збудження — збільшення енергії молекули: вона переходить із основного коливального стану  в деякий збуджений коливальний стан  так, що енергетична різниця між цими рівнями дорівнює енергії фотона.



Енергія поглинутого інфрачервоного випромінювання витрачається на збудження коливальних переходів для речовин у конденсованому стані. Для газів поглинання кванта ІЧ-випромінювання викликає коливальні й обертальні переходи.

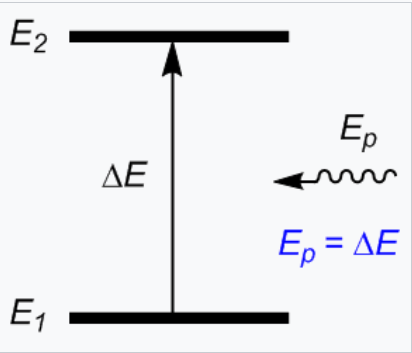


Рис. 1.1. Поглинання електромагнітного випромінювання

Інфрачервона спектроскопія займається головним чином вивченням молекулярних спектрів, оскільки в ИК-області розташована більшість коливальних і обертальних спектрів молекул. Дослідження будови молекул і інших об'єктів методами інфрачервоної спектроскопії має на увазі отримання відомостей про параметри молекулярних моделей і математично зводиться до рішення зворотних спектральних задач. Рішення таких задач здійснюється послідовним наближенням шуканих параметрів, розрахованих за допомогою спеціальної теорії спектральних кривих до експериментальних.

Інфрачервона спектроскопія дозволяє ідентифікувати просторові ізомери, вивчати всередині- і міжмолекулярні взаємодії, характер хімічних зв'язків, розподіл зарядів в молекулах, реєструвати короткоживучі (час життя до 106с) частинки, отримувати дані для обчислення термодинамічних функцій, встановлення форми нормальних коливань, розподіл коливальної енергії по мірах свободи, положення смуг в спектрах і їх інтенсивності. Розрахунки спектрів молекул, що містять до 100 атомів, в тому числі полімерів, виконуються з допомогою ЕОМ. При цьому необхідно знати характеристики молекулярних моделей, які знаходять рішенням відповідних зворотних спектральних задач або квантово-хімічними розрахунками. Тому інфрачервона спектроскопія як метод вивчення будови молекул набув найбільшого поширення в органічній хімії. У окремих випадках для газів в інфрачервоній області вдається спостерігати обертальну структуру коливальних смуг. Це дозволяє розраховувати дипольні моменти і геометричні параметри молекул, уточнювати силові постійні і т. д.

Інфрачервона спектроскопія має ряд переваг перед спектроскопією у видимій і ультрафіолетовій областях, оскільки дозволяє прослідити зміну всіх основних типів зв'язків в молекулах досліджуваних речовин. При використанні інфрачервоної спектроскопії для визначення якісного і кількісного складу природних сумішей не відбувається руйнування речовин, що дозволяє застосовувати їх для подальших досліджень.

1.2. Загальна характеристика методів інфрачервоної спектроскопії

Інфрачервона область відкрита в 1800 г англійським астрономом У. Гершелем, а в 1905 р американський фізик Кобленц опублікував великий огляд ІЧ спектрів багатьох класів органічних і неорганічних сполук.

В основі отримання ІЧ спектрів лежить пряме поглинання світла при проходженні через шар речовини. Знімають ІЧ спектри розчинів речовин, тонких плівок, паст, твердих речовин. Кожен спосіб має свою методику зняття ІЧ спектру.

Кювети для зняття ІЧ спектрів виготовляються з кристалічних галогенідів лужних металів, які не дозволяють працювати з водними, кислими і лужними розчинами. Для цієї мети існують кювети зі спеціальних матеріалів, наприклад з Ge, Si, - [CH2-CH2]n- і ін. Якщо досліджувана речовина добре розчиняється в органічних розчинниках, то його ІЧ спектр знімають у вигляді розчину. При дослідженні ІЧ спектрів розчинів речовин розчинник повинен бути, по можливості, інертний, не мати власних інтенсивних смуг в досліджуваній області. Сполуки, які погано розчинні (нерозчинні), вивчають у вигляді суспензії у вазеліновій олії або запресовують у пластинки з КBr.

Повний ІЧ спектр органічної сполуки лежить в діапазоні 400- 4000 см-1 . Діапазон лабораторних ІК спектрометрів від 100-3500см1; саме в цьому діапазоні поглинають більшість органічних молекул. Поглинання молекулою енергії в цьому діапазоні викликає зміну коливальних станів атомів, що входять до складу молекули, і обертальних станів молекул.

ІЧ спектроскопія грунтується на вивченні коливань одних атомів щодо інших. Важливу роль відіграють відносні коливання двох атомів, пов'язаних між собою хімічним зв'язком. Коливання великих частин молекули не відіграють важливої ролі.

Число і положення смуг у спектрі залежать від природи та кількості атомів у молекулі, її геометричної будови, симетрії, а також від потенціального поля внутрішньомолекулярних сил.

Інтенсивності смуг визначаються електронними властивостями молекул: дипольним моментом (μ), поляризованістю (α), а також їх змінами в процесі коливань.

Молекула може поглинати квант ІЧ випромінювання лише в тому випадку, якщо це приводить до зміни її дипольного моменту. ІЧ спектри здебільшого представляють в координатах , хвильове число зменшується зліва направо (рис. 1.2).

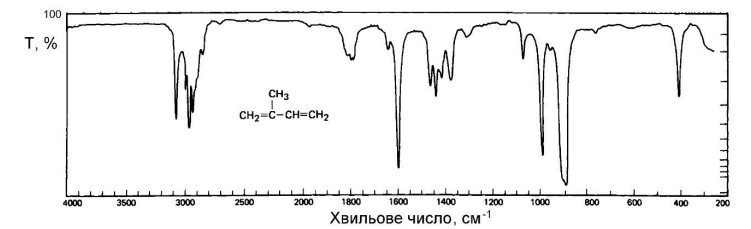


Рис. 1.2. ІЧ спектр ізопрену

2. МЕТОДИ МАШИННОГО РОЗПІЗНАВАННЯ

2.1. Машинне навчання та його типи

Машинне навчання (Machine learning, ML) – група методів в області штучного інтелекту, набір алгоритмів, які застосовують для створення машини, яка вчиться на власному досвіді. В якості навчання машина обробляє величезні масиви вхідних даних і знаходить в них закономірності. За вже звичним словом «алгоритм» ховається складна робота з його налаштування та навчання. Алгоритми адаптивно покращують свою ефективність, оскільки збільшуються дані, доступні для навчання. Тобто чим більше даних, тим точнішою буде наша модель і ефективнішим навчання.

Мета машинного навчання – частково або навіть повністю автоматизувати рішення різних складних аналітичних задач.

Тому в першу чергу машинне навчання покликане давати максимально точні прогнози на підставі вхідних даних, щоб людина могла приймати правильні рішення в своїй роботі. В результаті навчання машина може передбачати результат, запам'ятовувати його, при необхідності відтворювати, вибирати кращий з декількох варіантів.

Машинне навчання має три основні складові (Рис. 2.1):

а) дані – базова інформація, в яку входять будь-які вибірки даних, роботі з якими потрібно навчити систему;

б) ознаки – які саме характеристики і властивості повинна відстежувати система в результаті навчання;

в) алгоритм – вибір методу для вирішення поставленого завдання.

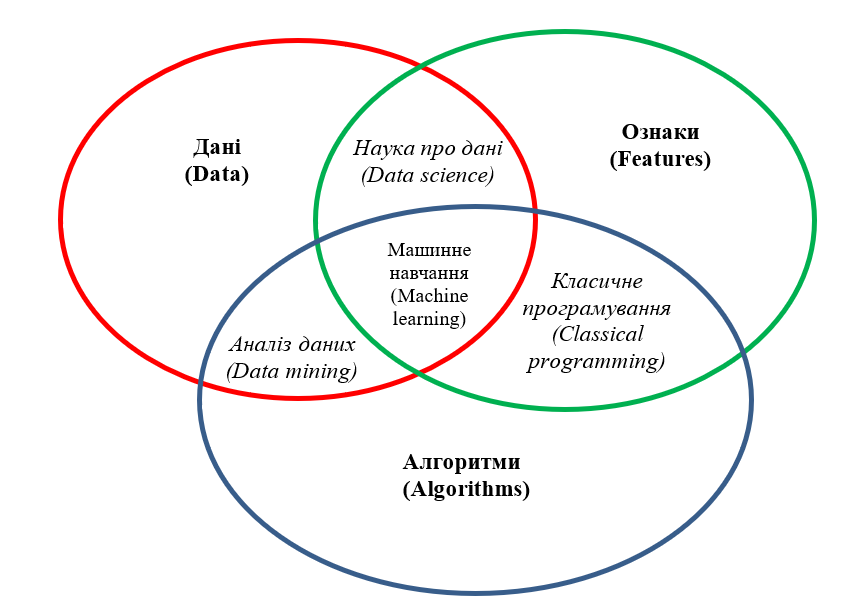


Рис. 2.1. Складові машинного навчання

2.2. Алгоритми розпізнавання паттернів

2.3 Застосування машинного розпізнавання в хімії

3. МЕТОДОЛОГІЯ ДОСЛІДЖЕННЯ

3.1. Збір та підготовка даних інфрачервоних спектрів

3.2. Вибір та налаштування алгоритмів машинного розпізнавання

3.3. Розробка програмного забезпечення для автоматизованої ідентифікації