Langages rationnels

Définition, propriétés,
Expressions régulières
Théorème de Kleene
Lemme de l'étoile

Réseaux de neurones

Auteur : Olivier Raynaud

Version de travail : rentrée 2020
Sensibilité :

Référence :

https://st.focusedcollection.com/13422768/i/650/focused_330408632-stock-photo-abstract-structure-neural-network-dark.jpg

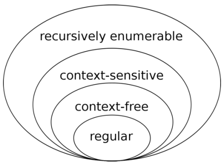
1

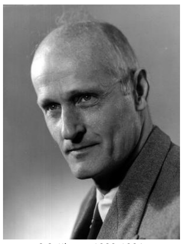
Langages rationnels

- *Langages rationnels* *(aussi appelés réguliers ou regular language en anglais)*
  - Tous langage fini peut être énuméré, pour les langages infinis une telle énumération exhaustive est impossible.

Définition intuition (langages rationnels) :
Ce sont les langages qui peuvent être définis, on dit aussi dénotés, par une expression régulière (au sens formel du terme).

Cette classe comprend ainsi tous les langages de taille finie.





S.C. Kleene 1909-1994

2

Langages rationnels

- *Opérations rationnelles*
  - Les opérations d'union, de concaténation et de fermeture de Kleene – appelée opération étoile - sont dites rationnelles.

Définition (langages rationnels noté Rat Σ^*) :
Soit Σ un alphabet, les langages rationnels sur Σ sont définis inductivement par :
(i) $\{\epsilon\}$ et $\emptyset$ sont des langages rationnels ;
(ii) $\forall a \in \Sigma$, $\{a\}$ est un langage rationnel ;
(iii) Si L_1 et L_2 sont des langages rationnels, alors $L_1 \cup L_2$, $L_1 L_2$ et L_1^* sont des langages rationnels

Note: tout langage fini est donc rationnel

Soient les langages $\{ab, c\}$ et $\{ba, c\}$ sur $\Sigma = \{a, b\}$, on a :
— $\{ab, c\} \cup \{ba, c\} = \{ab, c, ba\}$;
— $\{ab, c\}.\{ba, c\} = \{abba, abc, cb, cc\}$;
— $\{ab, c\}^* = \{\epsilon, ab, c, abab, abc, cab, cc, ...\}$;

3

Expressions rationnelles

❑ *Les expressions rationnelles définissent un système de formules*

➢ Un langage rationnel peut se décomposer sous la forme d'une formule finie, correspondant aux opérations rationnelles qui permettent de le construire.

Définition (expressions rationnelles) **:**
Soit Σ un alphabet, l'ensemble des expressions rationnelles sur Σ noté **RE** est défini inductivement par :

(i) ϵ et $\emptyset$ sont des expressions rationnelles ;
(ii) $\forall a \in \Sigma$, a est une expression rationnelle ;
(iii) Si e_1 et e_2 sont des expressions rationnelles, alors $(e_1 + e_2)$, $(e_1 e_2)$ et (e_1^*) sont des expressions rationnelles.

Exemple

$((((re)g)e)x)$ est une expression rationnelle par ii) et par iii).

4

Propriétés

Propriétés :
Soient **w** un mot et **r** une expression rationnelle, nous avons :

$\epsilon w = w = w\epsilon$
$\emptyset w = \emptyset = w\emptyset$
$\emptyset + r = r = r + \emptyset$
$\epsilon^* = \epsilon$
$\emptyset^* = \epsilon$
$(\epsilon + r)^* = r^*$

Propriétés :
Soient **r** une expression rationnelle et **a** et **b** des symboles,nous avons :

$(r^*)^* = r^*$
$r(r^*) = (r^*)r = r^+$
$(a^*b^*)^* = (a+b)^*$

5

Interprétation

❑ *Les expressions rationnelles*

➢ Une expression rationnelle permet d'exprimer, de déclarer les mots - chaînes de caractères - que l'on veux accepter.

Interprétation (expressions rationnelles) :
L'interprétation d'une expression est définie par les règles inductives suivantes :

(i) ϵ dénote le langage $\{\epsilon\}$ et $\emptyset$ dénote le langage vide ;
(ii) $\forall a \in \Sigma$, a dénote le langage $\{a\}$;
(iii) $(e_1 + e_2)$ dénote l'union des langages dénotés par e_1 et e_2 ;
(iv) $(e_1 e_2)$ dénote la concaténation des langages dénotés par e_1 et e_2 ;
(v) (e_1^*) dénote la fermeture de Kleene du langage dénoté par e_1.

Pour simplifier les notations un ordre de priorité sur les opérateurs est fixé.
Ainsi l'étoile (*) est l'opérateur le plus liant, puis la concaténation puis l'opérateur d'union.

$((((re)g)e)x)$ peut donc s'écrire $regex$.

6

Expression régulière et langage

Notation :
Soit **r** une expression rationnelle, on notera **L(r)** le langage dénoté par **r**.

Définir en langage naturel, les ensembles dénotés par les expressions suivantes :

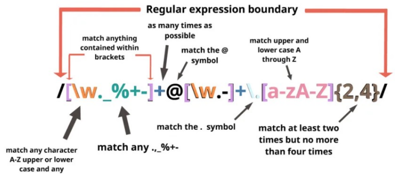
$r_1 = 00$
$r_2 = (0+1)^*$
$r_3 = (0+1)^*00(0+1)^*$
$r_4 = 0^4 1 0^4$

$L(r_1) =?$, $L(r_2) =?$ $L(r_3) =?$ et $L(r_4) =?$

7

Langage des adresses courriel

Définition du langage des adresses courriel



Exemples d'adresses valides :

- `Abc@example.com`
- `Abc@10.42.0.1`
- `Abc.123@example.com`
- `user+mailbox/department=shipping@example.com`
- `!#$%&'*+-/=?^_`.{|}~@example.com`

Exemples d'adresses non valides :

- `Abc.example.com`
  - Le caractère `@` manque.
- `Abc.@example.com`
  - Le caractère `.` est situé juste avant le caractère @
- `Abc..123@example.com`
  - Le caractère `.` apparaît deux fois de suite.

https://fr.wikipedia.org/wiki/Adresse_%C3%A9lectronique#Histoire

8

Expressions régulières et langages rationnels

Théorème (Equivalence entre expressions rationnelles et langages rationnels) **:**
Un langage est rationnel ***si et seulement si*** il est dénoté par une expression rationnelle.

Eléments de démonstration

------>
Confère les éléments de preuve donné plus loin dans l'exposé.

<-----
Confère les éléments de preuve donné plus loin dans l'exposé.

9

Equivalence

❑ *Relation d'équivalence entre expressions rationnelles*
- Un même langage rationnel peut être décrit par plusieurs expressions régulières.

Définition (équivalence) **:** Soient deux expressions rationnelles **r** et **s**,
On dit que **r** est **équivalent** à **s**, et on note **r~ s**, ***si*** **r** et **s** dénotent le même langage, autrement dit **si** L(r) = L(s).

Considérons le langage des mots sur $\Sigma = \{a, b\}$ contenant le facteur aa.
Donner une définition formelle de ce langage.

Soient les expressions rationnelles suivantes :
$r = (a+b)^*aa(a+b)^*$
$s = b^*(ab^+)^*aa(a+b)^*$

Les langages $L(r)$ et $L(s)$, équivalents, définissent formellement le langage des mots contenant le facteur aa.

10

Mise en pratique

❑ *Question*
- Définir l'ensemble des mots de taille quelconque contenant alternativement des 0 et des 1.

❑ *Réponse*
- 01*, (01)*
- (01)* + (10)*
- (01)* + (10)* + 1(01)* + 0(10)*

❑ *Réponse*
- $(\epsilon + 1)(01)^*(\epsilon + 0)$

❑ *Questions*
- Donner l'expression rationnelle dénotant le langage des mots sur $\Sigma = \{a, b, c\}$ contenant au moins un a et au moins un b.
- Donner l'expression rationnelle dénotant le langage des mots sur $\Sigma = \{0, 1\}$ contenant au moins une paire de 1 consécutifs.
- Donner l'expression rationnelle dénotant le langage des mots sur $\Sigma = \{0, 1\}$ pour lesquels chaque pair de 0 consécutifs apparait devant une paire de 1 consécutifs.

11

Théorème de Kleene

❑ *Théorème de Kleene.*

Définition (Langage reconnaissable par un AEF) **:**
On appelle **langage reconnaissable** sur un alphabet Σ tout langage qui peut être reconnu par un automate à états fini.
L'ensemble des langages reconnaissables est noté Rec Σ***.**

Théorème (équivalence entre les A.E.F. et les langages rationnels) **:**
Soit un alphabet fini Σ, il y a **égalité** entre l'ensemble des langages rationnels et l'ensemble des langages reconnaissables par un A.E.F.
Autrement dit nous avons :

$$\text{Rat } \Sigma^* = \text{Rec } \Sigma^*$$

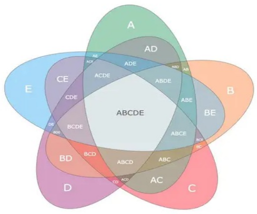
12

Propriétés de fermeture

❑ *Propriétés de fermeture*

➢ L'ensemble des langages rationnels est fermé (ou stable) pour plusieurs opérateurs.

Définition intuition (fermeture pour une opération) **:** On dit d'un ensemble E qu'il est **fermé** pour un opérateur donné, si cet opérateur appliqué aux éléments de E donne toujours un résultat appartenant à E.



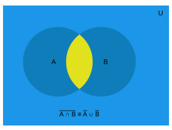
Un exemple de clôture dans le cas fini.

https://www.greelane.com/fr/science-technologie-math%c3%a9matiques/math/how-many-elements-in-the-power-set-3126439/

13

Propriétés de fermeture des langages rationnels

Propriétés (fermeture) **:**
Soient L_1 et L_2 des langages rationnels, alors $L_1 \cup L_2$, $L_1 L_2$, L_1^*, $L_1 \cap L_2$ et $\overline{L_1}$ sont des langages rationnels.



Eléments de démonstration

Les langages rationnels sont fermés par les opérateurs rationnels (cf. Définition)

Le complémentaire peut se démontrer avec élégance via le formalisme des A.E.F.

Enfin, avec les lois de De Morgan, nous avons $L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$ et nous pouvons donc conclure pour l'opérateur d'intersection.

14

Equivalence entre RE(Σ) et Rec Σ*

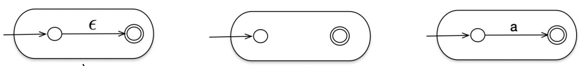
Théorème (expressions et langages) **:**
Tout langage rationnel dénoté par une expression rationnelle est aussi reconnu par un AEF.

Eléments de démonstration
------>

Soit L le langage reconnu par une expression rationnelle **r**. Montrons qu'il existe un AFN-e **A** tel que L(**r**)=L(A).

La preuve se fait par induction structurelle sur **r** en suivant la définition récursive des expression rationnelle donnée plus haut.

Base :



a) b) c)

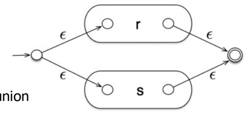
15

Equivalence entre RE(Σ) et Rec Σ*

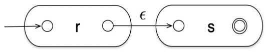
Théorème (expressions et langages) **:**
Tout langage rationnel dénoté par une expression rationnelle est aussi reconnu par un AEF.

Eléments de démonstration

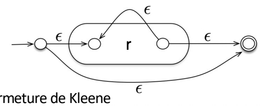
Phase d'induction:



a) L'union



b) La concaténation



c) La fermeture de Kleene

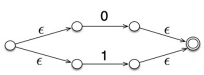
16

Exemple de construction d'automate

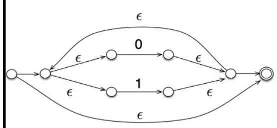
❑ *Construction d'un automate à partir d'une expression rationnelle :*

➢ Soit l'expression (0+1)* 1 (0+1)

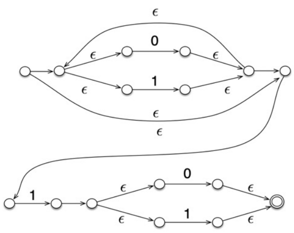
Eléments de construction



r_0 : 0+1



r_1 : (0+1)*



r_2 : (0+1)* 1 (0+1)

17

Grammaire régulière

❑ *Grammaire régulière*

Définition (Grammaire régulière) **:** Une grammaire G=(N,T,P,S) est dite **régulière** si toutes les règles de production dans P sont de la forme:

$$A \to \alpha B, \text{ avec } A \in N, \alpha \in T^*, B \in N \text{ ou } B = \epsilon.$$

Exemple: G1 : S -> aS | bS | a | b | ϵ.

Cette classe de grammaire est composée de 2 sous classes :

Les grammaires **régulières** (linéaires) **à droite**, où toutes les règles sont de la forme:

$$A \to \alpha B, \text{ avec } A \in N, \alpha \in T^*, B \in N \text{ ou } B = \epsilon.$$

Les grammaires **régulières** (linéaires) **à gauche**, où toutes les règles sont de la forme:

$$A \to B\alpha, \text{ avec } A \in N, \alpha \in T^*, B \in N \text{ ou } B = \epsilon.$$

18

Grammaires régulières et langages rationnels

Théorème (Equivalence entre grammaires régulières et langages rationneles) **:**
Un langage est rationnel ***si et seulement si*** il est engendré par une grammaire régulière à gauche ou par une grammaire régulière à droite.

Eléments de démonstration

------> Soit A un automate avec A=(Q,Σ, δ, q_0, F), nous construisons un grammaire G=(N,T,P,S) régulière de la façon suivante:
N coïncide avec Q, T coïncide avec Σ, S coïncide q_0 et pour toute transition de δ de la forme (q_i, α) -> q_j. on crée une règle de production dans P de la forme Q_i -> α Q_j

Exemple :

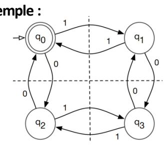


Diagramme de transitions de A

On obtient la grammaire G=(N,T, P, S) telle que :
- N = {Q_0,Q_1,Q_2,Q_3} ;
- T= {0,1}
- P :

  Q_0 -> 1Q_1, Q_0 -> 0Q_2 Q_0 -> ε
  Q_1 -> 1Q_0, Q_1 -> 0Q_3
  Q_2 -> 1Q_3, Q_2 -> 0Q_0
  Q_3 -> 1Q_2, Q_3 -> 0Q_1
- S = Q_0

19

Grammaires régulières et langages rationnels

Théorème (Equivalence entre grammaires régulières et langages rationneles) **:**
Un langage est rationnel ***si et seulement si*** il est engendré par une grammaire régulière à gauche ou par une grammaire régulière à droite.

Eléments de démonstration

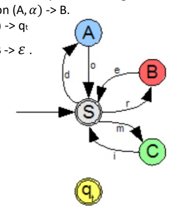
<-----
Pour tout grammaire régulière G=(N,T,P,S), il existe un AFN A=(Q,Σ, δ, q_0, F) tel que L(G) coïncide avec le langage reconnu par A.
L'ensemble des états Q vaut N u {q_t}, l'ensemble Σ vaut T, q0 est l'état S et pout toute règle de production de P de la forme A -> αB, on insère dans δ la transition (A, α) -> B.
Et pour les règles de la forme A -> α, on insère la transition (A, α) -> q_t

F est l'ensemble des symboles B de N tels qu'il existe une règle B -> ε .

Exemple

La grammaire suivante est une grammaire régulière à droite :

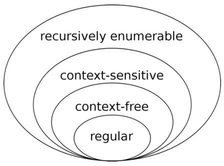
$$S \to dA \mid rB \mid mC \mid \epsilon$$
$$A \to oS$$
$$B \to eS$$
$$C \to iS$$



20

Frontière des langages rationnels

- *A la frontière des langages rationnels*
  - L'ensemble des langages rationnels est petit devant l'ensemble de tous les langages.



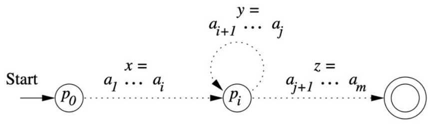
- *Question : comment montrer qu'un langage est rationnel ?*
- *Question : comment montrer qu'un langage n'est pas rationnel ?*

21

Lemme de l'étoile

Lemme de l'étoile (appelé aussi lemme de la pompe)

Théorème (lemme de la pompe pour les langages rationnels) **:**
Soit L un langage rationnel, alors il existe une constante n telle que pour tout mot w de L tel que $|w| \leq n$, le mot w peut être décomposé en trois facteurs, $w = xyz$, tels que :
— $y \neq \epsilon$;
— $|xy| \leq n$;
— Pour tout $k \geq 0$, le mot xy^kz appartient aussi à L.



22

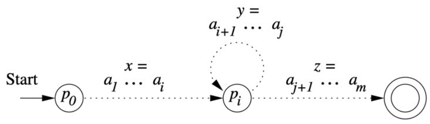
Lemme de l'étoile

Lemme de l'étoile (appelé aussi lemme de la pompe)

Element de démonstration :
Si un langage L est rationnel alors il existe un A.E.F. déterministe A minimum tel que L=L(A).
Soit n le nombre d'états de A.

Soit un mot w quelconque de L(A) tel que $w=a_1a_2...a_m$ avec m la taille de w supérieure à n;
Soit $p_0,p_1,...,p_k$ la liste des états par lesquels passe A à la lecture du mot w ;
Soit $p_i = \delta(q_0,a_1a_2...a_i)$ pour tout i dans [0,..m];

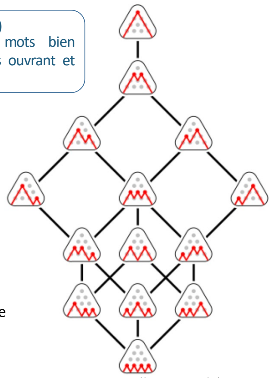
Alors puisque |w|>n, il existe i et j tel que $p_i = p_j$



23

Langage de Dyck

Définition (Langage des mots bien parenthésés)
Un langage de Dyck est l'ensemble de mots bien parenthésés sur un alphabet fini de symboles ouvrant et fermant.


https://stringfixer.com/fr/Dyck_language

Exemple :
Sur l'alphabet composé de la paire de parenthèses {'(', ')'} , le mot '(())()' est un mot bien parenthésé alors que le mot '()))()' ne l'est pas.

Propriété :
Si L est un langage de Dick sur {'(', ')'} **alors** L={w ∈ {(,)}* | $|w|_(= |w|_)$ et tout préfixe propre de w contient plus de '(' que de ')'};

24

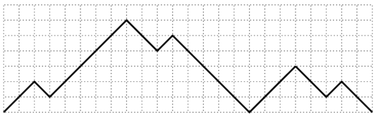
Contraposée du lemme de l'étoile

Contraposée du Lemme de l'étoile :
Soit L un langage, si pour toute constante n, il existe un mot w de L avec |w|>= n tel que pour toute décomposition de w en xyz avec |xy| <= n et |y|>=1, il existe i >= 0 tel que le mot xy^iz n'est pas dans L alors L n'est pas rationnel.

Exemple d'application : Langage de Dyck
Montrons que le langage L des mots bien parenthésés sur { (,) } n'est pas rationnel.

Soit la constante **n**, considérons w =$(^{n+1}\ )^{n+1}$. On a bien w ∈ L;
Toute décomposition de w en xyz avec |xy| ≤ n impose que x et y ne sont composés que de '('.
Or, pour par exemple i=2, le mot xy^2 z n'est pas dans L car ce mot possède plus de '(' que de ')';

En vertu de la contraposée
Du lemme de l'étoile nous
En déduisons que
L n'est pas rationnel;



http://www-igm.univ-mlv.fr/~jyt/L3_MPI4/mpi4_2020_Cours3.html

25

Pour résumer

- *Quelques points essentiels de l'exposé*
  - **Définitions :** Langage rationnel, expression rationnelle (regular expression) ;
  - **Théorème de Kleene :** équivalence entre l'ensemble des langages rationnels et l'ensemble des langages reconnus par un E.A.F.
  - **Lemme de l'étoile** (pumping lemma) pour démontrer qu'un langage n'est pas rationnel.

26

Bibliographie

- [HMU] **Introduction to Automata Theory, Langage and Computation**
  J.E. Hopcroft, R. Motwani, J.D. Ullman Edition Adison Wesley 2001;
- [DPPA10] **Logicomix**
  A. Doxiadis, C. Papadimitriou, A. Papadatos, A.D. Donna Edi. Vuibert 2010;
- [Berry 00] **Support de cours de Théorie des Langages**
  A. Berry – Université Clermont-Auvergne 2000 – 2016
- [Wikipedia] **Multiples références dans le texte.**

27