

La gestion des incertitudes dans les modèles à base de cas en estimation des coûts de logiciels

Ali Idri, Alain Abran

Liste des sujets

- Introduction
- La technique CBR en estimation des coûts
- Sources d'incertitude:
 - Erreurs dans le mesurage
 - Imprécisions dans l'estimation
- Version non-déterministe
- Discussion

Introduction

- Raisonement à base de cas:
 - Version simplifiée du raisonnement par analogie
 - Se base sur l'hypothèse que la résolution d'un problème consiste en l'accès à des informations mémorisées
 - Utilise des informations spécifiques à des problèmes déjà résolus (= des cas) plutôt que des informations générales sur le domaine

Introduction

- 4 étapes du CBR:
 - Retrouver les cas similaires au cas étudié
 - Utiliser les cas similaires retrouvés pour proposer une solution au problème
 - Réviser la solution proposée
 - Retenir les éléments de cette expérience (cas + solution) pour la résolution de problèmes futurs

Introduction

Plusieurs chercheurs en estimation du logiciel se sont basés sur l'affirmation suivante:

- *'similar software projects have similar costs'*
- Technique prometteuse, mais s'applique difficilement avec des variables linguistiques souvent floues et imprécises

Introduction

- Technique CBR : donne une seule valeur estimation
- Mais, même si des projets sont similaires, leurs coûts ne sont pas nécessairement similaires!
- Il faut donc identifier l'incertitude, et un intervalle de valeurs d'estimation relié à cette incertitude

Nouvelle approche: *Fuzzy Analogy*

- Identification des projets par des valeurs linguistiques traitées par des éléments de logique floue
- Évaluation de la similarité entre projets à partir de valeurs linguistiques
- Dans l'étape adaptation, utilisation d'une nouvelle stratégie pour déduire le coût à partir des coûts réels de projets similaires.

Sources d'incertitude dans les estimations

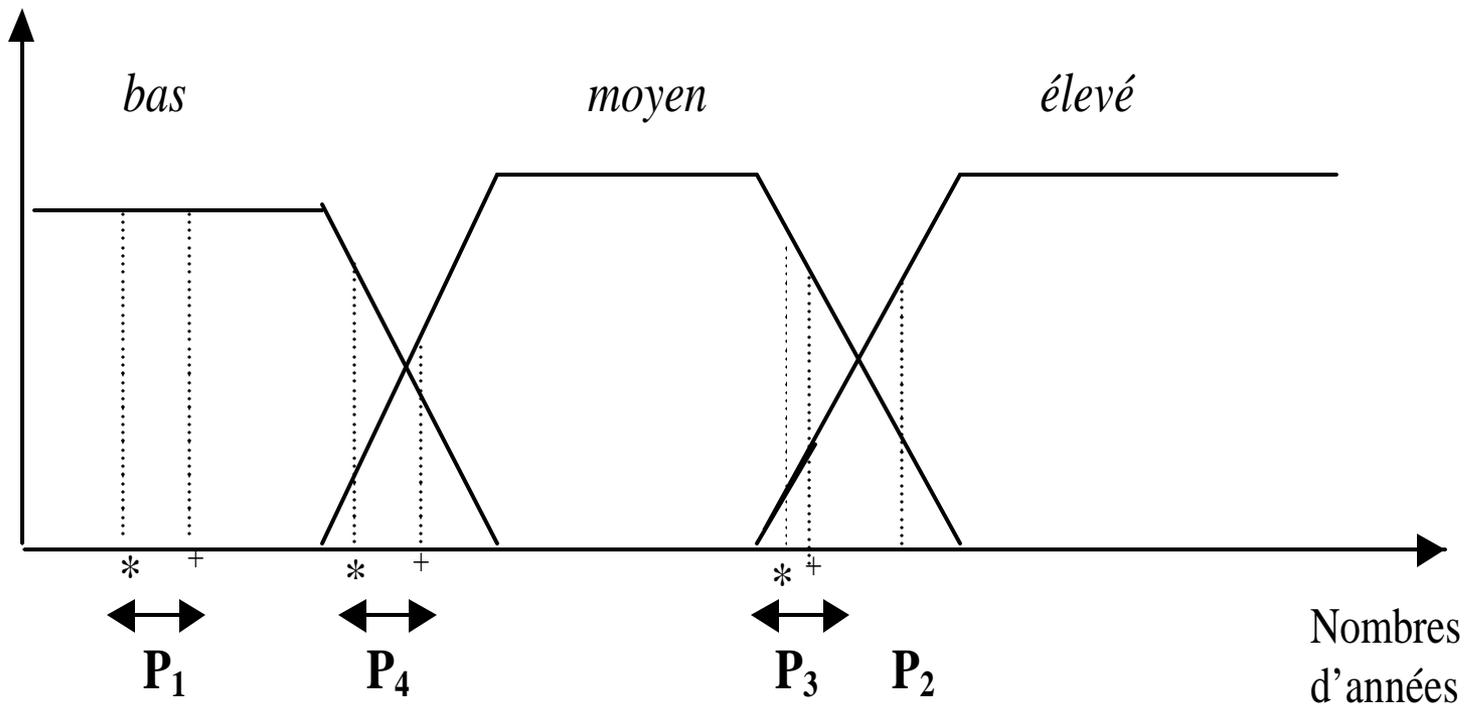
- 4 sources principales (Kitchenham & Linkman):
 - **Erreurs de mesurage des attributs affectant les coûts**
 - **Erreurs du modèle**
 - Erreurs relatives aux hypothèses faites sur le modèle
 - Erreurs relatives aux caractéristiques de l'environnement à partir duquel le modèle a été mis au point

Erreurs dans le mesurage des attributs

- Utilisation des valeurs linguistiques & numériques:
 - Faciles à comprendre
 - Permettent la tolérance à des imprécisions dans le processus de mesurage
 - Généralisent les valeurs numériques qui ne sont utilisées que si des informations précises sont disponibles

Attribut: Expérience des programmeurs

1



Exemples d'erreurs de mesurage

- P1 est à la position (*) au lieu de la position (+):
 - dans les deux cas P1 appartient à la valeur linguistique *bas* avec un degré d'appartenance égal à un.
 - Cette erreur de mesurage sur l'expérience du programmeur P1, n'affecte pas l'estimation du coût.
- P3 est à la position (*) au lieu de la position (+):
 - P3 appartient aux deux valeurs linguistiques *moyen* et *élevé* avec deux degrés d'appartenance différents.
 - Si nous évaluons la similarité individuelle de P3 et P2 par l'agrégation *max-min*, nous obtenons la même valeur dans les deux cas (P3 est dans la position (*) ou (+)).

Erreur de mesurage

- Avec des valeurs linguistiques et des fonction d'appartenance:
 - Bien des erreurs de mesure n'ont pas d'impact
 - Seulement certains cas ont un impact.

Erreur du modèle

- Incertitudes relatives aux imprécisions dans les estimation fournies par *Fuzzy Analogy*

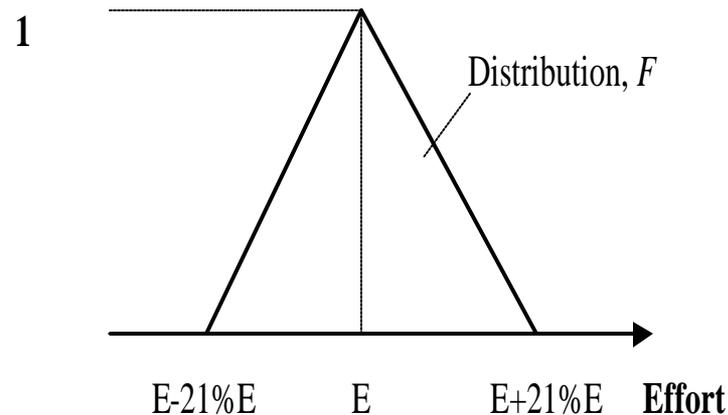
Précision d'un modèle d'estimation

Dépend de:

- Approche adoptée pour la mise au point du modèles:
 - Régression linéaire
 - Réseau de neurones
 - CBR
- Caractéristiques de la base de projets à partir de laquelle le modèle a été mis au point
- Caractéristiques de la base de projets pour valider le modèle

Critère d'évaluation de précision

- Précision du Fuzzy ANalogy avec COCOMO I :
MMRE = 21%
- D'où incertitude de + ou - 21%



Critique du MMRE comme évaluateur de précision

- Kitchenham & Likman:
 - MMRE accorde les mêmes degrés d'incertitude pour les sur-estimations que pour les sous-estimations
 - Proposent une distribution de probabilité (eg. Gamma)

Incertitudes en *Fuzzy Analogy*

« Similar projects have similar costs »

- Intuitivement correct, mais...
 - Projets logiciels peuvent ne pas être décrits de manière précise et exact
 - Les mesures de similarité entre les projets, ainsi qu'entre les coûts peuvent être non valides
 - Peut ne pas être vrai en pratique!!!

Déterministe ou non?

« Similar projects have similar costs »

- Si $d(P_i, P_j) \leq C(c_i, c_j)$, l'affirmation est déterministe pour le couple (P_i, P_j) . Cela signifie que la valeur de vérité de la proposition P_i et P_j sont similaires doit être inférieure ou égale à celle de la proposition c_i et c_j sont similaires.
- Si $d(P_i, P_j) > C(c_i, c_j)$, l'affirmation est non-déterministe pour le couple (P_i, P_j) .

Expérimentation avec COCOMO I

- Chercher tous les projets historiques P_i qui sont étroitement similaires à P , ensuite, déterminer pour chaque P_i , l'ensemble des valeurs similaires au coût de P_i avec un degré de similarité supérieur ou égal à $d(P, P_i)$, $E_i(P)$:

- $$E_i(P) = \{c / C(c, c_i) \geq d(P, P_i)\}$$

- où c_i est le coût de P_i . Les valeurs possibles au coût du P sont donc données par l'ensemble $E(P)$:

- $$E(P) = \bigcap_i E_i(P)$$

Expérimentation avec COCOMO I

- En appliquant cette procédure de la version déterministe à COCOMO'81, dans la plupart des cas, l'ensemble $E(P)$ est vide:
 - En effet, un projet P peut être similaire à plusieurs projets P_i mais les coûts des P_i sont totalement différents.
 - Par exemple, le projet P_9 de la base du COCOMO'81 est évaluée être similaire à P_{27} et P_{56} avec respectivement les degrés 0,30 et 0,19.
 - Cependant, il n'existe aucune valeur numérique qui est à la fois similaire, au coût de P_{27} (88 K\$) aussi bien qu'au coût de P_{56} (985 K\$).

Version non-déterministe

Reformulation:

- « Similar software projects have **possibly** similar costs »

– La modélisation de cette règle de possibilité, en utilisant la conjonction pour l'implication, donne l'ensemble flou de toutes les valeurs possibles au coût du P:

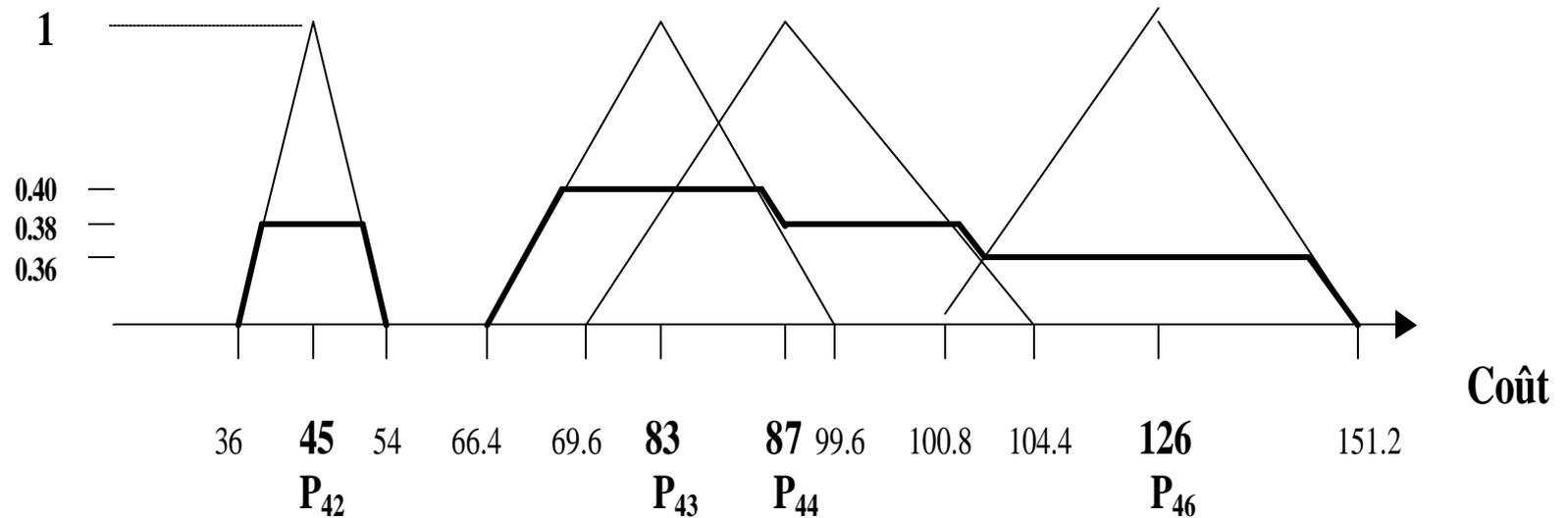
$$\mathbf{P}_{\text{cost}(P_i)}(c) = \min(d(P, P_i), C - R_{\text{cost}(P_i)}(c))$$

Version non déterministe pour un ensemble de N projets

- Chaque projet P_i génère un ensemble de valeurs possibles au coût du P.
 - Ces ensembles flous sont combinés par l'opérateur *max* (opération de disjonction) afin d'obtenir l'ensemble flou C_p contenant toutes les valeurs possibles du coût du projet P:

$$C_P(c) = \max_i (\mathbf{p}_{cost(P_i)}(c))$$

Exemple: Projet P45



Exemple: Projet P45

- P45 est similaire aux projets P42, P43, P44, et P46 avec des degrés respectivement égaux à 0,38, 0,40, 0,38 et 0,36.
- La distribution de possibilités est définie seulement pour les valeurs réelles similaires à au moins une des quatre valeurs représentant les coûts de P42, P43, P44, et P46;
- Pour les autres valeurs, la distribution de possibilités est inconnue.

Estimation

- La détermination d'une estimation numérique au coût d'un projet P à partir de sa distribution de possibilités utilise, en général, des techniques de *defuzzification* telles que la technique du centre de gravité et la technique *mean-of-maxima*.
 - Ces techniques ne sont pas forcément appliquées à tout l'ensemble flou dont la fonction d'appartenance est la distribution de possibilités du coût. Elles peuvent être appliquées seulement à une partie de cet ensemble flou.
 - Par exemple, dans le cas du projet P45, il est fort probable que son coût réel soit dans l'intervalle [66,4, 152,2]. Ainsi, l'application de la technique du centre de gravité à cet intervalle génère une estimation égale à 100 K\$ au coût du P45;
 - Le coût réel de P45 est 106 K\$.

Sommaire

- Une nouvelle version floue de l'estimation des coûts par analogie: *Fuzzy Analogy*
 - Tolérance aux imprécisions tout au long du processus d'estimation:
 - Dans la description des projets par des valeurs linguistiques
 - Dans l'évaluation de la similarité entre projets
 - Dans l'utilisation de la valeur linguistique *étroitement similaire* pour la déduction du coût

Questions

?