Prediction on Rank Promotion based on Career Sequence of Government Official

Jyi-Shane Liu & Long-Ping Lai
Department of Computer Science
National Chengchi University

Background

- Digitization of historical documents
- Subject databases in humanities and social sciences
- Text and data analysis techniques
  - Text mining, data mining
- A new frontier for interdisciplinary research
  - Domain knowledge – problem, result
  - Computational techniques

Purpose

- To investigate the pattern and factors of government official's rank promotion

- Data – job change records of government official
- Computational techniques – data mining
- Problem definition
  - Binary classification: to predict whether or not an official will be promoted to a target rank within a specified timeframe based on his/her current career sequence

Data Domain

- Rank promotion summarizes job achievement of government official's career

- Taiwan's rank system – fourteen levels in three tiers
  - 委任 – Rank level: 1 – 5
  - 薦任 – Rank level: 6 – 9
  - 簡任 – Rank level: 10 – 14
  - Rank information only available after 1988

An Example of Career Sequence

單位	職等	職稱	時間
臺灣高等法院花蓮分院檢察署	十	檢察官	840811
臺灣高等法院檢察署	十	檢察官	850401
臺灣高等法院檢察署	十一	檢察官	860805
法務部	十一	副司長	870702
臺灣苗栗地方法院檢察署	十二	檢察長	891020
臺灣高等法院檢察署	十二	檢察官	891201
福建金門地方法院檢察署	十二	檢察長	891212
臺灣基隆地方法院檢察署	十三	檢察長	900808
臺灣苗栗地方法院檢察署	十三	檢察長	900815
臺灣基隆地方法院檢察署	十四	檢察長	921030
臺灣桃園地方法院檢察署	十四	檢察長	921104
法務部	十四	司長	940517

(Representation: 10: 2, 11: 4, 12: 1, 13: 3, 14: 6)

More Examples of Career Sequence

姓名	職等陞遷序列屬性資料				
王增瑜	10:3	11:3	12:3	13:3	14:6
王憲義	10:5	11:3	12:2	13:3	14:8
白文漳	10:2	11:4	12:3	13:2	14:8
何明楨	10:6	11:2	12:3	13:3	14:3
石素梅	10:4	11:5	12:3	13:10	14:0
吳萬得	10:3	11:4	12:3	13:11	14:0
劉科	10:7	11:9	12:0	13:0	14:0
方芷絮	10:8	11:9	12:0	13:0	14:0
林燕山	10:4	11:13	12:0	13:0	14:0
吳癸受	10:4	11:12	12:0	13:0	14:0

Duty Type

- Chief
  - X長, 主任
- Deputy
  - 副XX
- Staff
- Representation: T (years in total), A (years as chief), B (years as deputy)
- Ex.: (10T:2, 10A:0, 10B:0, 11T:4, 11A:0, 11B:3, 12T:1, 12A:1, 12B:0, 13T:3, 13A:3, 13B:0, 14T:6, 14A:6, 14B:0)

Binary Classification of Rank Promotion

- Target rank: 13 and 14
  - The top two ranks
  - Chief of important government units (局, 處, 署, 司) or deputy minister
- Classify a career sequence to one of two classes (YES or NO: promotion to a higher rank will occur in the specified time interval)
  - Given a sequence of ranks 10-13, predict its promotion to rank 14
  - Given a sequence of ranks 10-12, predict its promotion to rank 13

Time Intervals of Observation and Prediction (I)

- Requirement
  - Full sequence above rank 10
  - Balance between positive and negative cases
- Inspection of available data
  - Approximately 15 years from rank 10 to rank 14
  - Positive cases (less than 15 years): 146
  - Negative cases (more than 15 years): 176
  - Approximately 12 years from rank 10 to rank 13
  - Positive cases (less than 12 years): 221
  - Negative cases (more than 12 years): 233

Time Intervals of Observation and Prediction (II)

- Two sets of sequence data

- 15-year sequence (for rank 14)
  - (observation, prediction): (12, 3)
  - (observation, prediction): (10, 5)

- 12-year sequence (for rank 13)
  - (observation, prediction): (9, 3)
  - (observation, prediction): (8, 4)

Data Mining Techniques

- Decision Tree algorithm
  - Classification based on hierarchical division based on attribute values
- Support Vector Machine
  - Classification based on statistical learning theory
- Training and testing on classification model

Experimental Design (I)

- Four subsets of experiments
  - Rank 14: (12, 3)
  - Rank 14: (10, 5)
  - Rank 13: (9, 3)
  - Rank 13: (8, 4)
- Parameters
  - Duty type information
  - Sequence scope for training: full, 2/3 front, 2/3 back

Experimental Design (II)

- Training and validation
  - 30% bootstrap, 50% bootstrap, 10-fold CV
- Evaluation index

Class	Predicted	
True	Yes	No
Yes	TP：True Positive	FN：False Negative (TYPE I ERROR)
No	FP：False Positive (TYPE Ⅱ ERROR)	TN：True Negative

  - Error rate: (FN+FP)/(TP+FN+FP+TN)

Experimental Results (I)

- Error rates in training phase
  - Support vector machine: 15%~21%
  - Decision tree: 15%~21%
- Error rates in testing phase
  - Support vector machine: 18%~23%
  - Decision tree: 20~25%
- Error rate increase
  - Support vector machine: 1%~7% (absolute diff.)
  - Decision tree: 1%~11% (absolute diff.)
- Support vector machine provides more stable performance

Experimental Results (II)

- More training cases result in somewhat better classification but not significant

- Duty type information (with or without)
  - Support vector machine: 2%~10% (absolute diff.)
  - Decision tree: 1%~6% (absolute diff.)
  - Duty type information is helpful

Experimental Results (III)

- Sequence scope effects
  - 2/3 (back) > full > 2/3 (front)
  - Difference between best and worst
    - Support vector machine: 10%~16% (absolute diff.)
    - Decision tree: 5%~12% (absolute diff.)
- Best performance
  - Use 2/3 (back) sequence and duty type information
- Data mining
  - Use key attributes and reduce noises

Experimental Results (IV)

- FN errors >>> FP errors
  - Worst FN: 77%~92% (2/3 front for rank 14), but FP 8%~2%
  - Too early to rule out later promotion
- FN/FP ratio
  - Worst in rank 14 prediction than in rank 13
  - Rank 14: special consideration from 特任官
  - Rank 13: personnel system stability and safeguard
- FP errors
  - No more than 10%, as low as 2%
  - System stability/safeguard for good job performance

Summary

- Support Vector Machine
  - Best error rate: 18%
  - Worst error rate: 40%
- Decision Tree
  - Best error rate: 19%
  - Worst error rate: 40%
- Best prediction performance
  - With duty type information + sequence scope: 2/3 back

Conclusion

- Prediction on rank promotion within specified time interval
  - Positive relation with rank-year sequence and duty type information
  - Approximately 77%~82% correct prediction
- Rank-year sequence is a good aggregate indicator of job performance and can be used to predict future promotion
- Much to be explored on humanity and social science data