

Active Contact System

Data science for contact management

Andrea Brunello ¹ Enrico Marzano ²

¹Ph.D. student at Università degli Studi di Udine

²CIO at Gap srl



Introduction

Agent performance

Written notes
analysis

Feature extraction

Clustering notes

Classifying notes

Call classification

Anomalous agents
detection

Conclusions

Introduction

Agent performance

Written notes analysis

Feature extraction

Clustering notes

Classifying notes

Call classification

Anomalous agents detection

Conclusions

Tracking the performance of agents is a primary issue in contact centers, as it allows, for example:

- ▶ the best match to be taken between service and agent;
- ▶ the recognition of unsatisfactory agent behaviours, due for example to a lack of proper training;
- ▶ the prediction of future trends, based on the history of observations.

Agent performance assessment at Gap

ACS

Brunello, Marzano

Introduction

Agent performance

Written notes
analysis

Feature extraction

Clustering notes

Classifying notes

Call classification

Anomalous agents
detection

Conclusions

- ▶ Agent evaluation is typically conducted by human supervisors;
- ▶ Gap srl employs also a custom analytic function to this end, leveraging:
 - ▶ **technical data**, for example:
 - ▶ mean conversation period;
 - ▶ mean post-call period;
 - ▶ **service data**, for example:
 - ▶ INB, written notes over phone sessions;
 - ▶ INB, call results;
 - ▶ OUTB, hourly revenue (positives over working time).
- ▶ Given an agent, the function outputs a score, i.e., a real number $\in [0, 1]$, which can be used to compile a ranking.

Other than *quantitative* values, also *qualitative* aspects are important to define the profile of an agent:

- ▶ **inbound calls:** analysis of written notes;
- ▶ **outbound calls:** automatic call result tagging.

Idea:

- ▶ how often / in which way does an agent record notes regarding an inbound call?
- ▶ compare single agent behaviour with service average values.

How to evaluate written notes?

- ▶ extract summarizing features from the text;
- ▶ identify groups of similar notes;
- ▶ devise a methodology to assign a generic new note to one of the groups.

Feature extraction from plain text

For each note, calculate (*R* script, *openNLP* library):

- ▶ number of words and characters;
- ▶ *Gulpease* readability index value;
- ▶ fraction of articles over words;
- ▶ fraction of conjunctions over words;
- ▶ fraction of verbs over words;
- ▶ fraction of adjectives over words;
- ▶ fraction of adverbs over words;
- ▶ fraction of prepositions over words;
- ▶ fraction of quantifiers over words;
- ▶ fraction of (pro)nouns over words;
- ▶ fraction of numeric codes over words;
- ▶ fraction of proper nouns over words;
- ▶ fraction of words/abbrs. found in Italian dictionary;
- ▶ fraction of words found in **service-specific** domain;
- ▶ fraction of unrecognized words.

Identify groups of similar notes

ACS

Brunello, Marzano

Introduction

Agent performance

Written notes
analysis

Feature extraction

Clustering notes

Classifying notes

Call classification

Anomalous agents
detection

Conclusions

- ▶ Random sampling of 1000 notes;
- ▶ application of a clustering algorithm to the selected notes (*E-M* algorithm);
- ▶ 6 clusters emerged:
 - ▶ articulated notes;
 - ▶ non-articulated notes;
 - ▶ abbreviated notes;
 - ▶ domain-specific notes;
 - ▶ nonsense notes;
 - ▶ hybrid notes.

Classify a new note

- ▶ Attach a new feature to each of the clustered notes:
cluster label
- ▶ apply a decision tree-inducting algorithm (*J48*) to the dataset, with the goal of predicting the label (94.7% accuracy);
- ▶ the tree can then be used to classify new notes.

```
riconosciuti_abbr_su_parole <= 0.142857
|  riconosciuti_dominio_su_parole <= 0.133333
|  |  preposizioni_su_parole <= 0
|  |  |  non_riconosciuti_su_parole <= 0.157895
|  |  |  |  congiunzioni_su_parole <= 0.025
|  |  |  |  |  articoli_su_parole <= 0.071429: non_articulated_notes
|  |  |  |  |  articoli_su_parole > 0.071429: articulated_notes
|  |  |  |  |  congiunzioni_su_parole > 0.025: articulated_notes
|  |  |  non_riconosciuti_su_parole > 0.157895
|  |  |  |  non_riconosciuti_su_parole <= 0.333333
|  |  |  |  |  articoli_su_parole <= 0.083333: hybrid_notes
|  |  |  |  |  articoli_su_parole > 0.083333: articulated_notes
|  |  |  |  non_riconosciuti_su_parole > 0.333333: non_sense_notes
|  |  preposizioni_su_parole > 0
|  |  |  indice_gulp <= 129.833333: articulated_notes
|  |  |  indice_gulp > 129.833333
|  |  |  |  non_riconosciuti_su_parole <= 0.0625: non_articulated_notes
|  |  |  |  non_riconosciuti_su_parole > 0.0625: hybrid_notes
```


Example

Introduction

Agent performance

Written notes analysis

Feature extraction

Clustering notes

Classifying notes

Call classification

Anomalous agents
detection

Conclusions

	valore text	gruppo_nota text
1	info voltura	hybrid
2	invio del f24 unico	articulated
3	informazioni per appunt sub e comunica dati catastali	articulated
4	info posizione pagamenti mensa scolastica	hybrid
5	NON RISPONDE	non-articulated
6	Info	abbreviated
7	VIA [REDACTED] MQ 37 C'è SCRITTO 43 BOLLETTAZIONE SBAGLIATA. DEVE PASSARE AGLI SPORTELLI PER RETTIFICA DI METRATURA CON PIANINA SCALA 1:100. RIFERISCO. C	articulated
8	SIGNORA CHIAMA PER SAPERE SE È STATA APPLICATA LA DETRAZIONE DI 25 euro per figlio sul calcolo	articulated
9	la signora aveva chiamato il 23/05 per una verifica posizione per la TARES: ha un locale come	articulated
10	chiede quanto deve pagare per la tassa. Parlati con esatto: deve pagare 61 euro.	articulated
11	info boll	abbreviated
12	rimborso ud	non-articulated
13	tasi	domain-specific
14	INFO GENERICHE IMU, TASI	domain-specific
15	info su avv sosp	hybrid
16	chiede se può rateizzare l'importo da versare per la mensa. Riferito che deve fare richiesta	articulated
17	invio copia boll	hybrid
18	chiede il saldo mensa. Riferito che abbiamo problemi tecnici tecnici al server	articulated

- ▶ Track distribution of note classes for each service;
- ▶ compare with agent-service distributions.

Is it possible to automatically classify a phone session, starting from all recorded information regarding it?

- ▶ *outbound* calls follow a predefined script;
- ▶ a workflow has to be respected by the agent;
- ▶ the different phases have typical durations;
- ▶ for example, an outbound survey campaign.

How to exploit such a capability?

- ▶ compare call results marked by agents with those suggested by the automatic classification;
- ▶ anomalous agents detection, to complement the performance module.

Sometimes, agents mark phone sessions with outcomes that are different with respect to what really happened during the call:

- ▶ because of a mistake;
- ▶ intentionally, for example simulating surveys which did not really take place.

We can detect such cases, evaluating, for each call, the durations of:

- ▶ dialling phase;
- ▶ conversation phase;
- ▶ post-call phase.

Given the outbound survey service of interest, we may group call session results into 5 categories:

- ▶ *engaged or non-existent number;*
- ▶ *answer-machine or fax;*
- ▶ *no answer;*
- ▶ *talked, but no survey;*
- ▶ *survey made.*

General approach:

- ▶ generation of a decision tree capable of predicting the result category of a phone session, basing on dialling, conversation and post-call times.

As a training set, we want to consider only correctly tagged sessions:

- ▶ we consider only logs generated by agents which are thought to be reliable;
- ▶ further deletion of tuples affected by *class noise*, and not by *attribute noise*. For example:
 - ▶ *surveys* with < 40 sec conversation;
 - ▶ *answer-machine* or *fax* without conversation;
 - ▶ talked, but no survey with < 5 sec conversation;
 - ▶ *engaged* or *non-existent* with conversation > 0 ?
⇒ **NO!**

We obtain a tree which tries to predict the most likely result category for a phone session, given the three time values ($> 93\%$ accuracy).

Resulting decision tree

```
tempo_di_conversazione <= 7
| tempo_di_conversazione <= 0
| | tempo_di_dial <= 30
| | | tempo_di_dial <= 11: occupato_o_inesistente
| | | tempo_di_dial > 11
| | | | tempo_di_dial <= 14: occupato_o_inesistente
| | | | tempo_di_dial > 14: nessuna_risposta
| | tempo_di_dial > 30: nessuna_risposta
| tempo_di_conversazione > 0
| | tempo_di_postcall <= 1
| | | tempo_di_dial <= 29: segreteria_o_fax
| | | tempo_di_dial > 29
| | | | tempo_di_conversazione <= 1: nessuna_risposta
| | | | tempo_di_conversazione > 1: segreteria_o_fax
| | tempo_di_postcall > 1
| | | tempo_di_conversazione <= 4: segreteria_o_fax
| | | tempo_di_conversazione > 4: parlato_negativo
tempo_di_conversazione > 7
| tempo_di_conversazione <= 76
| | tempo_di_conversazione <= 11
| | | tempo_di_postcall <= 1
| | | | tempo_di_conversazione <= 9
| | | | | tempo_di_dial <= 22
| | | | | | tempo_di_conversazione <= 8: segreteria_o_fax
| | | | | | tempo_di_conversazione > 8: parlato_negativo
| | | | | | tempo_di_dial > 22: segreteria_o_fax
| | | | tempo_di_conversazione > 9: parlato_negativo
| | | tempo_di_postcall > 1: parlato_negativo
| | tempo_di_conversazione > 11: parlato_negativo
| tempo_di_conversazione > 76
| | tempo_di_conversazione <= 87
| | | tempo_di_postcall <= 0: parlato_negativo
| | | tempo_di_postcall > 0: sondaggio_effettuato
| | tempo_di_conversazione > 87: sondaggio_effettuato
```

⇒ Exploit the tree to derive “problematic” agents or sessions, to be further investigated.

Introduction

Agent performance

Written notes analysis

Feature extraction

Clustering notes

Classifying notes

Call classification

Anomalous agents detection

Conclusions

Introduction

Agent performance

Written notes analysis

Feature extraction

Clustering notes

Classifying notes

Call classification

Anomalous agents
detection

Conclusions

- ▶ Importance of agent performance evaluation in the contact center domain;
- ▶ agent performance evaluation at Gap srl;
- ▶ case studies:
 - ▶ inbound written notes classification;
 - ▶ outbound call classification.