

Instituto Politécnico de Lisboa  
Instituto Superior de Engenharia de Lisboa

# **Otimização da Escolha de Matrículas Corretas num Sistema de Portagem**

Hugo Emanuel Lopes Fernandes

**Licenciatura em Matemática Aplicada à Tecnologia e à Empresa**

2020

Instituto Politécnico de Lisboa  
Instituto Superior de Engenharia de Lisboa



**A-to-Be**

Powered by  **Briso**

## **Otimização da Escolha de Matrículas Corretas num Sistema de Portagem**

Hugo Emanuel Lopes Fernandes

**Licenciatura em Matemática Aplicada à Tecnologia e à Empresa**

**Responsável:**

Mestre Lara Cordes Lima Trigueiro de Moura

**Orientadores:**

Mestre Luís Pedro Pereira da Silva

Doutora Sandra Isabel Cardoso Gaspar Martins

Doutor Sérgio Paulo Fino de Sousa Lopes

Janeiro 2020



# Resumo

O projeto R100, que engloba este estágio, é uma atualização ao sistema atual de pagamento automático de portagens; visa tanto aumentar a eficácia do sistema, como também adaptar o sistema a um maior número de mercados. Neste estágio, trabalhou-se à volta do algoritmo que o R100 utiliza para calcular o nível de confiança da leitura das matrículas dos veículos que passam nas portagens utilizadas no estudo.

Inicialmente, realizou-se uma fase de adaptação aos dados com o objetivo de entender a estrutura de organização e o comportamento dos dados; seguidamente realizou-se uma análise dos erros, paralelamente com a otimização de algumas variáveis do algoritmo.

Seguiu-se uma abordagem analítica diferente da utilizada até ao momento, com o intuito de eliminar alguns defeitos no algoritmo em uso. Com base nessa análise, fez-se uma reformulação do algoritmo e o estudo do desempenho desta nova fórmula.

# Agradecimentos

Em primeiro lugar gostaria de agradecer aos meus orientadores, Mestre Luís Pedro Silva, Doutora Sandra Gaspar Martins e Doutor Sérgio Lopes, pela disponibilidade e auxílio oferecido no decorrer de todo o estágio; foram um apoio fundamental na realização do mesmo.

Agradeço ao Mestre Francisco Mourão e ao Bruno Basílio pela partilha de ideias e críticas feitas perante os resultados apresentados durante o decorrer do estágio.

Agradeço à A-to-Be pela oportunidade, apoio e condições fornecidas.

Agradeço aos meus colegas da empresa por me terem recebido bem e pelo tempo passado em conjunto.

Por fim quero agradecer à minha namorada, família e amigos por todo o apoio, força e motivação dada ao longo deste período.

# Índice

<b>Lista de Figuras</b>	<b>v</b>
<b>Lista de Tabelas</b>	<b>vi</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Introdução . . . . .	1
1.2 Matrículas . . . . .	1
1.3 Projeto R100 . . . . .	1
1.4 Tratamento de Dados . . . . .	4
1.4.1 Agregação . . . . .	4
1.4.2 Filtragem . . . . .	4
1.4.3 Computação . . . . .	4
1.4.4 Classificação . . . . .	5
<b>2 Primeira Fórmula</b>	<b>7</b>
2.1 Tratamento de Dados com a Primeira Fórmula . . . . .	7
2.1.1 Exemplo . . . . .	8
2.2 Primeira Execução . . . . .	13
2.3 Atualizações do Sistema e Otimização das Variáveis . . . . .	15
2.3.1 Atualização no Motor MKW . . . . .	15
2.3.2 Atualização no Motor NLABS . . . . .	16
2.3.3 Atualização das Fotografias de Overview . . . . .	17
2.3.4 Junção das Atualizações . . . . .	18
2.4 Análise de Matching Levels de Erros . . . . .	19
2.5 Análise de Erros . . . . .	20
2.6 Estudo dos Thresholds . . . . .	20
2.7 Introdução da Regra de Sintaxe – Simulação Vs Algoritmo . . . . .	21
<b>3 Nova Fórmula</b>	<b>23</b>
3.1 Reformulação . . . . .	23
3.2 Funções de Confiança . . . . .	24
3.2.1 Funções de Confiança dos Motores . . . . .	24
3.2.1.1 ARH . . . . .	24
3.2.1.2 MKW . . . . .	24
3.2.1.3 NLABS . . . . .	25
3.2.2 Funções de Confiança dos Matching Factors . . . . .	26

## ÍNDICE

3.2.2.1	Photo Matching Factor . . . . .	26
3.2.2.2	Engine Matching Factor . . . . .	27
3.2.2.3	Front-Rear Matching Factor . . . . .	28
3.3	Primeira Execução . . . . .	28
3.4	Estudo dos Thresholds . . . . .	29
3.5	Alternativas . . . . .	29
3.6	Análise de Erros . . . . .	30
3.6.1	Ajuste de Erros . . . . .	31
3.7	Resultados da análise de erros . . . . .	31
<b>4</b>	<b>Conclusão</b>	<b>33</b>
4.1	Conclusão . . . . .	33
4.2	Desenvolvimento do estudo . . . . .	33

# Lista de Figuras

1.1	Sistema de Portagens Atual . . . . .	2
1.2	Sistema de Portagens do Projeto R100 . . . . .	3
1.3	Diagrama do Algoritmo Plate Selector . . . . .	4
1.4	Fluxograma de Decisão da Escolha da Matrícula Frontal . . . . .	6
2.1	Excerto do Ficheiro de Ficheiro de Output com Cálculos . . . . .	12
2.2	Histograma dos Resultados da 1ª Execução do Algoritmo . . . . .	13
2.3	Histograma dos Resultados para a Execução dos Valores Otimizados da Atualização do Motor MKW . . . . .	16
2.4	Histograma dos Resultados para a Execução dos Valores Otimizados da Atualização do Motor NLABS . . . . .	17
2.5	Histograma dos Resultados para a Execução dos Valores Otimizados da Atualização das Fotografias de Overview . . . . .	18
2.6	Histograma dos Resultados para a Execução dos Valores Otimizados da Junção das Atualizações . . . . .	19
2.7	Histograma dos Resultados com differentCharsThreshold=1 e differentCharsMinConfidenceThreshold=76 . . . . .	22
2.8	Histograma dos Resultados com differentCharsThreshold=2 e differentCharsMinConfidenceThreshold=67 . . . . .	22
3.1	Gráfico da Função de Reconhecimentos do motor ARH . . . . .	24
3.2	Gráfico da Função de Reconhecimentos do motor MKW . . . . .	25
3.3	Gráfico da Função de Reconhecimentos do motor NLABS . . . . .	26
3.4	Gráfico da Função de Photo Matching Factor . . . . .	27
3.5	Gráfico da Função de Engine Matching Factor . . . . .	27
3.6	Gráfico da Função de Front-Rear Matching Factor . . . . .	28
3.7	Histograma com os Resultados da Primeira Execução c/ nova Fórmula . . . . .	29
3.8	Histograma com os Resultados da Execução com Translação de Matrículas Estrangeiras . . . . .	31
3.9	Histograma com os Resultados da Execução com Translação de Matrículas Estrangeiras e Regra de Sintaxe Atualizada . . . . .	32

# Lista de Tabelas

2.1	Tabela de Reconhecimentos da Transação 42301241 . . . . .	9
2.2	Tabela dos Agregados Filtrados da Transação 42301241 . . . . .	9
2.3	Tabela de Cálculo do Nível de Confiança dos Agregados de transação 42301241 . . . . .	11
2.4	Tabela do Número de Transações com Front-Rear Matching Factor Nulo por Classe de Veículo . . . . .	14
2.5	Tabela com Número de Ocorrência e o Motivo . . . . .	14
2.6	Tabela com Número de Erros de Matching Level Nulos por Atualização . . . . .	19
2.7	Tabela com o Número de Melhor Overall/Frontal Certo/Errado . . . . .	20
2.8	Tabela de Variação dos Thresholds e Número de erros . . . . .	21
2.9	Tabela de Variação dos Thresholds e Número de erros Utilizando Regra de Sintaxe . . . . .	21
3.1	Tabela com o Número de Erros para cada Variação . . . . .	30



# Capítulo 1

## Introdução

### 1.1 Introdução

A A-to-Be é uma empresa portuguesa, pertencente ao grupo Brisa Inovação e Tecnologia, responsável pela criação, desenvolvimento e fornecimento de soluções dirigidas para o mercado de serviços de mobilidade internacional.

Este estágio foi realizado na sede da A-to-Be, em Lisboa, decorrendo de 16 de setembro de 2019 a 15 de fevereiro de 2020.

Numa realidade cada vez mais avançada tecnologicamente, qualidades como a rapidez, eficácia e comodidade são o que leva à procura de um produto; como tal, o serviço de pagamento eletrónico de portagens, em Portugal denominado por Via Verde, é bastante utilizado de modo a evitar perder tempo tanto em filas como no pagamento das portagens. Uma das ideias da A-to-Be é melhorar este sistema de pagamento eletrónico de portagens, com o intuito de aumentar a eficácia do mesmo e também criar a possibilidade da adaptação deste produto a um maior número de mercados: assim surge o projeto R100.

### 1.2 Matrículas

As matrículas são chapas exteriores dos automóveis cujo objetivo é identificar o veículo. Atualmente podemos encontrar em circulação no nosso país as sequências XX-00-00, 00-00-XX e 00-XX-00, em que X representa uma carácter do alfabeto, excluindo as letras K, Y e W, e 0 representa um número de 0 a 9. Existem também outras sequências mas são reservadas a algumas organizações, ou tipos de veículos. Existem três tipos de veículos que podem passar numa portagem: motas (veículos com apenas matrícula na sua traseira), veículos de classe 4 (veículos cuja matrícula da frente identifica o veículo, e a de trás identifica o reboque transportado), e restantes veículos (onde tanto a matrícula frontal como a traseira são iguais). Para o nosso estudo é importante referenciar as sequências associadas a reboques; estas podem ter as sequências entre X-0 e X-000000, ou entre XX-0 a XX-000000, onde X ou XX representam o código do serviço distrital ou regional onde o reboque foi registado.

### 1.3 Projeto R100

Atualmente o sistema de pagamento automático de portagens realiza uma leitura de um identificador, fornecido pela empresa Via Verde (empresa responsável por este tipo de portagens), que está associado a um cliente e consequentemente a uma conta bancária do mesmo. Esta leitura realiza-se quando um

## 1. INTRODUÇÃO

utilizador atravessa a portagem com o seu veículo, acto que vamos denominar de transação ao longo deste relatório. No entanto, no caso de existir uma falha na leitura do identificador, independentemente do motivo, existe um sistema que tira uma fotografia frontal, uma fotografia traseira, e grava a passagem do veículo. Como foi referido anteriormente, o projecto R100 visa melhorar este sistema, atingindo uma eficácia de 100% no reconhecimento de placas de matrículas, no entanto, caso existam erros, é desejável que se encontrem abaixo de um certo nível de confiança.

O sistema atual, que se encontra representado na Figura 1.1 tem a seguinte estrutura:

- Uma antena
- Dois sensores magnéticos
- Uma câmara frontal
- Uma câmara traseira
- Uma câmara de Overview

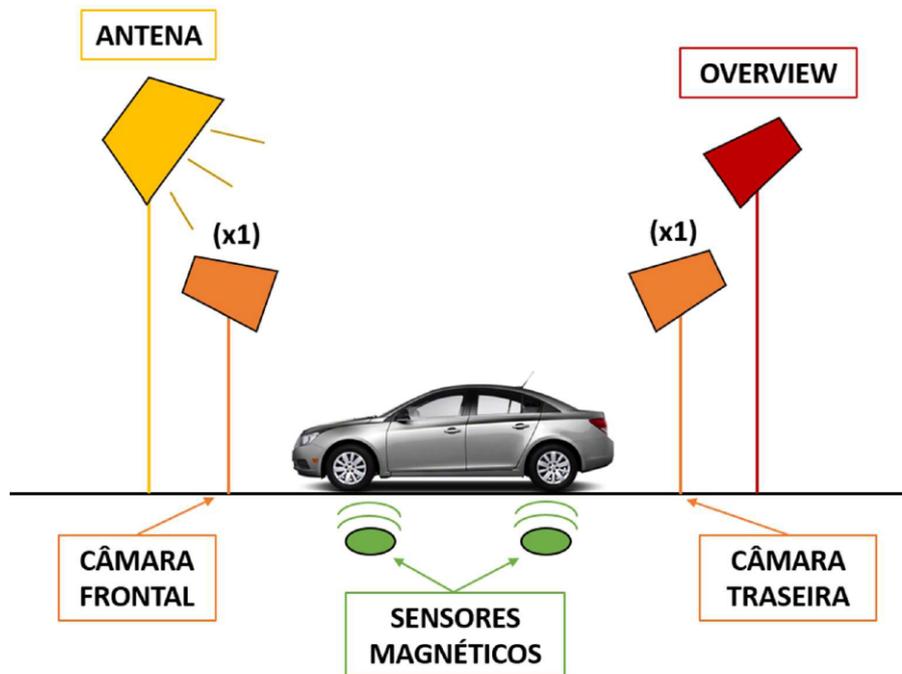


Figura 1.1: Sistema de Portagens Atual

Quando ocorre uma passagem na portagem, os sensores magnéticos captam a presença do veículo. Ao atravessar pelo primeiro sensor, a câmara frontal tira uma fotografia frontal ao veículo, e ao atravessar pelo outro sensor, a câmara traseira tira uma fotografia traseira. Durante este processo, a antena faz a leitura do identificador e a câmara de Overview filma toda a passagem.

Este é o sistema de que dispomos atualmente nas portagens Via Verde; no entanto, existem alguns defeitos no mesmo: problemas de iluminação, reflexão da luz solar nas placas de matrícula, mau estado das mesmas, ou até mau funcionamento das câmaras, são os principais fatores que podem impedir que a matrícula seja identificada. Situações como estas e a ausência do identificador, foram os principais motivos pelos quais se achou necessidade de criar o projecto R100.

Com a introdução deste projeto, o sistema que se encontra representado na Figura 1.2 passa a ter a seguinte estrutura:

- Um laser
- Duas câmaras frontais
- Duas câmaras traseiras
- Uma câmara de overview

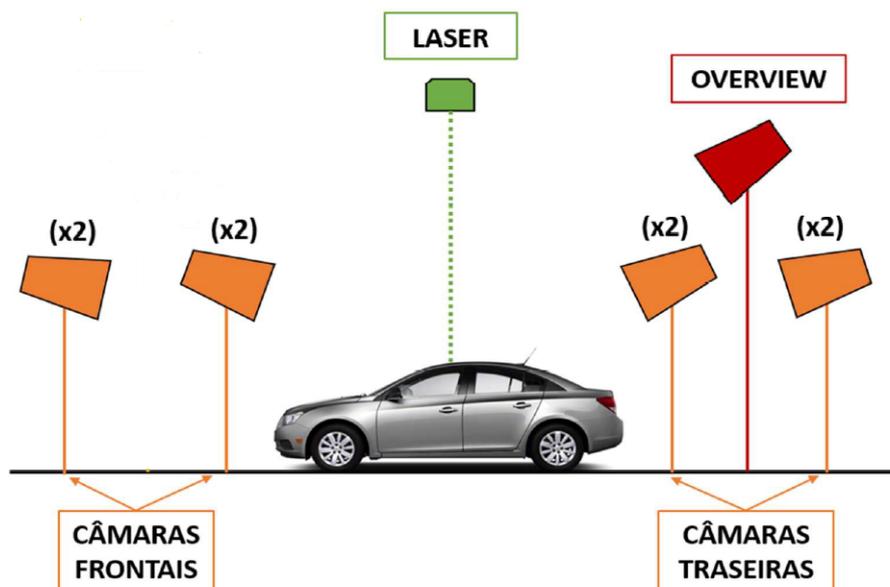


Figura 1.2: Sistema de Portagens do Projeto R100

O processo da travessia da portagem é semelhante ao anterior: ao atravessar a portagem, quando o veículo interceta o laser, cada câmara frontal tira duas fotografias frontais com aberturas de lente diferente, e ao terminar o contacto com o laser, cada câmara traseira tira também duas fotografias traseiras com aberturas de lente diferente, e câmara de overview tira uma fotografia. o comportamento da câmara de overview é semelhante, só que ensta versão esta além da filmage, tira também uma fotografia.

Após a captação das fotografias, as quatro fotografias frontais, as quatro traseiras e a de overview, são sujeitas a uma análise de OCR( Optical Character Recognition), análise esta que será realizada por três motores de reconhecimento de caracteres, produtos de três empresas diferentes: a ARH, a Makewise (MKW) e a NeuroLabs (NLABS).

Ao analisar uma fotografia, cada motor devolve um output com uma possível leitura de matrícula, denominada por reconhecimento, e o nível de confiança deste mesmo resultado. Ao passar todas as fotografias das transações do conjunto de estudo por todos os motores, juntamos todos os outputs num ficheiro, ficheiro esse que serve de input para o nosso algoritmo.

Para se poder avaliar o sistema de reconhecimento é essencial um conjunto de transações com manual review, uma revisão manual em que operadores analisam as fotografias das transações e identificam a matrícula correta. Como tal, utilizaram-se todas as transações do dia 14 de agosto de 2019 e os respectivos manual reviews; embora seja apenas um dia, este conjunto de dados contém transações de locais

## 1. INTRODUÇÃO

diferentes, de todas as horas do dia e conseqüentemente de diferentes ângulos em relação à reflexão da luz solar, o que torna a amostra significativa para ser estudada.

### 1.4 Tratamento de Dados

Após obtermos os reconhecimentos para uma transação, o passo seguinte é o tratamento desses dados e, para tal, é utilizado o algoritmo Plate Selector. O Plate Selector é um algoritmo de seleção de matrículas, criado para este projeto, que através de uma fórmula calcula o nível de confiança final tendo em consideração alguns aspectos adicionais – e não apenas o nível de confiança dos reconhecimentos. A criação deste algoritmo foi baseada na experiência e intuição dos membros da A-to-Be que fundaram este projeto. Este algoritmo está dividido em quatro etapas: a agregação, a filtragem, a computação, e a classificação, como se pode ver na Figura 1.3

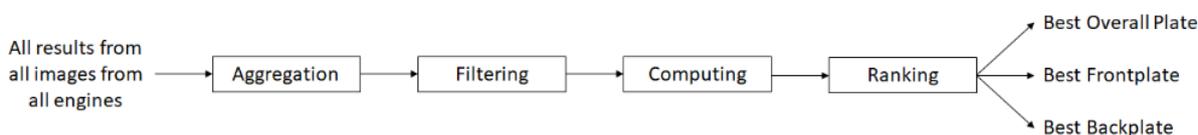


Figura 1.3: Diagrama do Algoritmo Plate Selector

#### 1.4.1 Agregação

A agregação é a etapa inicial, responsável pela reestruturação e agrupamento dos reconhecimentos. Todos os reconhecimentos são analisados e, por cada reconhecimento diferente, é criado um agregado com a matrícula obtida nesse reconhecimento. Nesta etapa é também calculado o nível de confiança médio de cada agregado, fazendo a média dos níveis de confiança dos reconhecimentos do mesmo.

#### 1.4.2 Filtragem

A filtragem é a etapa onde, para cada agregado, vamos analisar se a matrícula do agregado teve origem numa foto frontal, traseira, ou ambas; teremos uma abordagem para cada caso, criando uma variável dataType, e para cada agregado vamos atribuir o seu valor conforme o caso que se verificar durante a filtragem. Olhando apenas para as fotografias frontais, procuram-se os agregados que tenham sido identificados nestas fotografias e define-se o seu dataType como frontal. Repete-se o processo para as fotografias traseiras e aos agregados encontrados atribui-se o dataType traseiro. Ao olhar para todo o conjunto das fotografias, procura-se por agregados originados tanto de fotografias frontais como traseiras. Estes agregados serão iguais aos agregados frontais e traseiros cujo reconhecimento seja o mesmo. Caso não sejam encontrados, repetem-se os agregados originados por fotografias frontais e os agregados originados pelas traseiras. Aos agregados obtidos por uma destas duas situações, definimos o seu dataType como overall.

#### 1.4.3 Computação

Este algoritmo tem como um dos seus princípios o aumento dos níveis de confiança de matrículas que verifiquem certas ocorrências; para tal foram criados três boosts, um para cada tipo de ocorrência: Photo Matching, Engine Matching e Front-Rear Matching.

O Photo Matching Boost tem como objetivo aumentar o nível de confiança de agregados que tenham sido identificados em fotografias diferentes, pressupondo que quanto maior for o número de fotografias para um certo reconhecimento, maior deverá ser a probabilidade deste estar correto.

O Engine Matching Boost beneficia os níveis de confiança de agregados cujas matrículas foram identificadas por vários motores, assegurando que quanto maior o número de motores de um reconhecimento, maior será a probabilidade deste ser o certo.

O Front-Rear Matching Boost aumenta o nível de confiança de agregados cujos reconhecimentos provêm de fotografias frontais e traseiras; pelas mesmas analogias dos outros boosts, quanto maior o número de pares de fotografias frontais e traseiras, maior deverá ser o nível de confiança.

A computação é a etapa do algoritmo onde é utilizada uma fórmula para o cálculo do nível de confiança de cada agregado, relacionando estes com os boosts anteriormente referidos.

### 1.4.4 Classificação

A classificação é a etapa responsável pela escolha da matrícula vencedora. Vai olhar para os agregados de cada dataType e escolher, para cada um, o agregado com maior nível de confiança final. Em caso de empate, o algoritmo escolhe um aleatoriamente. O algoritmo olha para as matrículas de dataType overall e vai escolher o agregado com maior nível de confiança que retorna então como matrícula final lida para a transação. No entanto isto pode ser problemático caso o veículo da transação seja de classe 4, pois as matrículas frontal e traseira são diferentes; como tal, os reconhecimentos frontais serão diferentes dos traseiros, o que resulta na inexistência de agregados do tipo overall com fotos frontais e traseiras. Isto pode ser um problema pois, caso o agregado final escolhido seja referente à matrícula traseira, a matrícula atribuída à transação estará errada.

Para evitar este problema, decidiu-se criar dois critérios que, caso sejam verificados, levam à escolha do agregado com dataType frontal com maior nível de confiança. Definiram-se duas variáveis, o differentCharsThreshold e o differentCharsMinConfidenceThreshold, que serão referidas como thresholds. Caso o número de caracteres diferentes entre o melhor agregado do tipo overall e o melhor agregado do tipo frontal seja maior que o valor definido pelo differentCharsThreshold, e esse agregado frontal tenha um nível de confiança maior que o definido pelo differentCharsMinConfidenceThreshold, o algoritmo escolhe como matrícula da transação o agregado frontal com maior nível de confiança.

Considerando que mOverall representa a melhor matrícula overall, mFrontal representa a melhor matrícula frontal,  $\text{differentCharsThreshold} = \text{minChar}$  e  $\text{differentCharsMinConfidenceThreshold} = \text{min-FrontConf}$ , a Figura 1.4 representa este processo.

## 1. INTRODUÇÃO

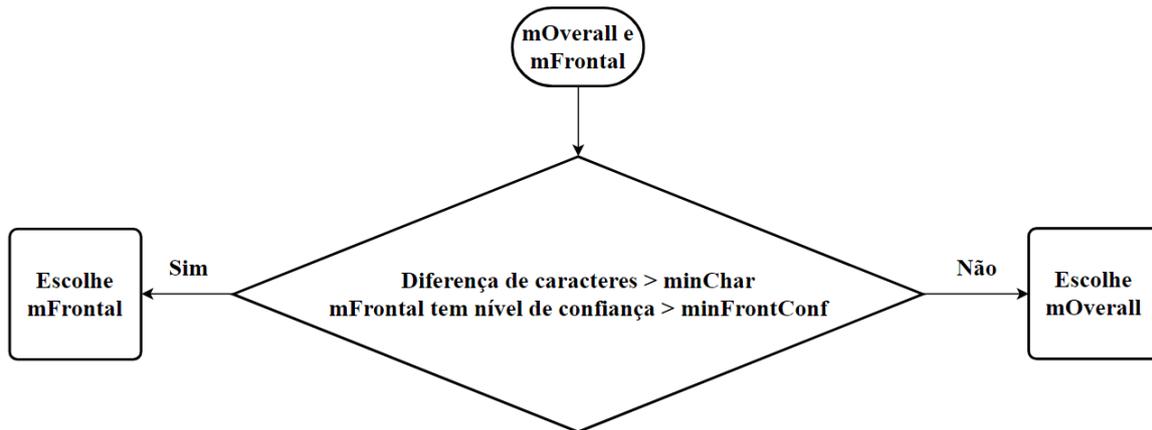


Figura 1.4: Fluxograma de Decisão da Escolha da Matrícula Frontal

Até este ponto, o algoritmo só estava a trabalhar com as quatro fotos frontais e as 4 traseiras; no entanto, ainda temos a fotografia de overview para acrescentar. Optou-se por utilizar esta fotografia, em casos onde o nível de confiança da matrícula final está abaixo de um certo valor, neste caso 50; volta-se a correr o algoritmo, mas agora incluindo a fotografia de overview com o objetivo de aumentar as probabilidades da matrícula selecionada para a transação estar correta.

## Capítulo 2

# Primeira Fórmula

Utilizando a fórmula original do projeto, o objetivo é otimizar o desempenho do algoritmo de maneira a diminuir ao máximo o número de erros – analisando os erros, otimizando as variáveis e introduzindo alternativas para a escolha da matrícula correta.

### 2.1 Tratamento de Dados com a Primeira Fórmula

Foi introduzido o algoritmo Plate Selector e as suas etapas do seu funcionamento. Tal como foi visto previamente, a computação é a etapa onde se calcula o nível de confiança dos agregados e é a única etapa dependendo da fórmula utilizada. A primeira fórmula consiste na soma de três parcelas, em que cada uma inclui cada um dos boosts.

A primeira parcela, denominada por Weighed Average, tem como objetivo conjugar o valor do nível de confiança médio com o peso de cada motor, onde cada motor vai tomar um valor percentual e a soma do peso dos três motores tem de ser 100%; é também a parcela onde se inclui o PhotoMatchingBoost e é representada pela fórmula 2.1

$$\sum([engineAvgConfidence + (nrEngineRecognitions - 1) * photoMatchingFactor] * engineWeight) \quad (2.1)$$

onde:

- *engineAvgConfidence* é a média aritmética dos níveis de confiança de um motor num reconhecimento;
- *nrEngineRecognitions* é o número de reconhecimentos de um motor num agregado;
- *photoMatchingFactor* é um parâmetro que é definido e representa o nosso Photo Matching Boost;
- *engineWeight* é o peso de cada motor; cada peso tem valores compreendidos entre 0 e 1, e a soma de todos os pesos tem de ser 1;

A segunda parcela, denominada por Engine Matching Level, tem como objetivo aumentar o valor do nível de confiança final quão maior for o número de motores com aquele reconhecimento, em relação ao número total de reconhecimentos. Se a matrícula for reconhecida apenas por um motor, esta parcela deverá ser nula. Esta parcela é representada pela fórmula 2.2:

## 2. PRIMEIRA FÓRMULA

$$\frac{engineMatchingFactor}{totNrRecognitionsAllAggregates} * \sum nrEngineRecognitions \quad (2.2)$$

onde:

- *engineMatchingFactor* é um parâmetro que é definido e representa o nosso Engine Matching Boost;
- *nrEngineRecognitions* é o número de reconhecimentos de um motor num agregado;
- *totNrRecognitionsAllAggregates* é o número total de reconhecimentos de todos os motores em todos os agregados de uma transação;

A terceira parcela, denominada por Front-Rear Matching Level, tem como objetivo aumentar o nível de confiança final quão maior for o número de pares de fotos frontais e traseiras, número este que é o menor valor de fotos frontais ou traseiras, pelo que esta parcela é representada pela fórmula 2.3:

$$\min(totNrFrontRecognitions, totNrBackRecognitions) * frontRearMatchingFactor \quad (2.3)$$

onde:

- *frontRearMatchingFactor* é um parâmetro que é definido e representa o nosso Front-Rear Matching Boost;
- *totNrFrontRecognitions* é o número total de reconhecimentos frontais de todos os motores num agregado;
- *totNrBackRecognitions* é o número total de reconhecimentos traseiros de todos os motores num agregado;

Não existe nenhum limite para o valor dos fatores utilizados no cálculo dos níveis de confiança. Portanto, ao utilizarmos um conjunto de valores relativamente elevados, o valor final para o nível de confiança de um agregado pode passar os 100%; assim sendo, no etapa da classificação utilizamos o majorante 100 para tal nível de confiança, após a escolha final do agregado.

### 2.1.1 Exemplo

Será apresentado um exemplo da execução do algoritmo para uma transação. Considerando que a transação em estudo é a número 42301241 e que:

- Photo Matching Factor toma o valor 12
- Front-Rear Matching Factor toma o valor 7
- Engine MatchingFactor toma o valor 12
- motor da ARH tem um peso de 0.34
- motor da MKW tem um peso de 0.33

## 2.1 Tratamento de Dados com a Primeira Fórmula

- motor da NLABS tem um peso de 0.33

Começando pela agregação, os reconhecimentos obtidos para esta transação estão representados na Tabela 2.1

Tabela 2.1: Tabela de Reconhecimentos da Transação 42301241

Tipo de Câmara	Motor	Matrícula Lida	Nível de Confiança
FRONTAL	ARH_CARMEN	PT15IX08	45
FRONTAL	NLABS_VPAR	PT51IX08	94
FRONTAL	ARH_CARMEN	PT51IX08	72
TRASEIRA	MKW_PLATE_VISION	PT51IX08	80
TRASEIRA	MKW_PLATE_VISION	PT51IX08	100
FRONTAL	MKW_PLATE_VISION	PT51IX08	100
FRONTAL	MKW_PLATE_VISION	PT51IX08	100
TRASEIRA	MKW_PLATE_VISION	PT51IX08	99
TRASEIRA	MKW_PLATE_VISION	PT51IX08	77
FRONTAL	MKW_PLATE_VISION	PT51IX08	93
FRONTAL	MKW_PLATE_VISION	PT51IX08	100
OVERVIEW	MKW_PLATE_VISION	PT51IX08	100

Ignorando a fotografia de overview, existem 11 reconhecimentos, sendo que 10 identificam a matrícula PT51IX08 e outro a matrícula PT15IX08; logo, os agregados existentes são PT51IX08 e PT15IX08.

Passando para a filtragem, começa-se pelo agregado PT15IX08; pela Figura 1.1, pode-se ver que o seu único reconhecimento provém da uma fotografia frontal e foi realizado pelo motor da ARH. Já o agregado PT51IX08 foi reconhecido tanto em fotografias frontais como traseiras, ignorando a de overview. Tanto o motor da NLABS como o da ARH apenas fizeram 1 reconhecimento frontal cada, enquanto que o motor da MKW fez 4 reconhecimentos frontais e 4 traseiros. Olhando apenas para as fotografias frontais, regista-se a média dos níveis de confiança dos reconhecimentos de cada motor para cada um dos agregados. Repete-se este processo olhando apenas para as fotografias traseiras, e por fim olhando para ambos os tipos de fotografias, registando os resultados na Tabela 2.2

Tabela 2.2: Tabela dos Agregados Filtrados da Transação 42301241

dataType	Matrícula	Nível de Confiança Médio	Motor
Frontal	PT15IX08	45	ARH
	PT51IX08	94	NLABS
		72	ARH
		98.25	MKW
Traseira	PT51IX08	89	MKW
Overall	PT51IX08	94	NLABS
		72	ARH
		93.625	MKW
	PT15IX08	45	ARH

A última linha da tabela vai ser ignorada a partir deste momento, pois enquanto que o agregado

## 2. PRIMEIRA FÓRMULA

PT51IX08 tem origem em fotografias frontais e traseiras, o agregado PT15IX08 apenas tem origem em fotografias frontais.

No processo de computação vamos calcular o nível de confiança para o agregado PT51IX08 do tipo overall. Começando pelo cálculo do valor da parcela Weigthed Average utilizando a fórmula 2.1 e os valores da Tabela 2.2:

Para o motor ARH:

$$\begin{aligned} & [engineAvgConfidence + (nrEngineRecognitions - 1) * photoMatchingFactor] * engineWeight = \\ & = [72 + (1 - 1) * 12] * 0.34 = 24.48 \end{aligned}$$

Para o motor MKW:

$$\begin{aligned} & [engineAvgConfidence + (nrEngineRecognitions - 1) * photoMatchingFactor] * engineWeight = \\ & = [93.625 + (8 - 1) * 12] * 0.33 = 58.61625 \end{aligned}$$

Para o motor NLABS:

$$\begin{aligned} & [engineAvgConfidence + (nrEngineRecognitions - 1) * photoMatchingFactor] * engineWeight = \\ & = [94 + (1 - 1) * 12] * 0.33 = 31.02 \end{aligned}$$

Logo:

$$WeigthedAverage = 24.48 + 58.61625 + 31.02$$

Procedendo agora ao cálculo da parcela Engine Matching Level, calculado através da fórmula 2.2, o motor da ARH e o motor da NLABS apenas obtiveram 1 reconhecimento neste agregado, enquanto que o motor da MKW obteve 8 reconhecimentos; logo:

$$\begin{aligned} & \frac{engineMatchingFactor}{totNrRecognitionsAllAggregates} * \sum nrEngineRecognitions = \\ & = \frac{12}{11} * (1 + 8 + 1) \approx 10.9090 = EngineMatchingLevel \end{aligned}$$

Quanto ao cálculo da parcela Front-Rear Matching Level, calculada através da fórmula 2.3, relembra-se que para este agregado existem 6 reconhecimentos frontais e 4 traseiros, assim sendo:

$$\begin{aligned} & \min(totNrFrontRecognitions, totNrBackRecognitions) * frontRearMatchingFactor \\ & = \min(6, 4) * 7 = 4 * 7 = 28 = Front - RearMatchingLevel \end{aligned}$$

Repete-se este processo para todos os agregados e obtém-se a tabela 2.3:

## 2.1 Tratamento de Dados com a Primeira Fórmula

Tabela 2.3: Tabela de Cálculo do Nível de Confiança dos Agregados de transação 42301241

<b>dataType</b>	<b>Matrícula</b>	<b>ConfMed</b>	<b>Motor</b>	<b>wAvg</b>	<b>EngineML</b>	<b>FrontRearML</b>	<b>ConfAgreg</b>
Frontal	PT15IX08	45	ARH	15.3	0	0	15.3
	PT51IX08	94	NLABS	99.8025	10.28571	0	110.08821
		72	ARH				
		98.25	MKW				
Traseira	PT51IX08	89	MKW	41.25	0	0	41.25
Overall	PT51IX08	94	NLABS	114.11625	10.90909	28	153.02534
		72	ARH				
		93.625	MKW				

onde ConfMed representa o nível médio de confiança de um agregado, wAvg representa a parcela Weighed Average, EngineML representa a parcela Engine Matching Level, FrontRearML representa a parcela Front-Rear Matching Level, e ConfAgreg representa o nível de confiança final do agregado.

Tendo os valores dos níveis de confiança dos agregados, podemos proceder à classificação. Para cada dataType seleciona-se o agregado com maior nível de confiança; neste caso, os melhores agregados frontal, traseiro e overall são PT51IX08. Assim sendo, não existe diferença de caracteres entre o melhor overall e o melhor frontal; logo, a matrícula será a melhor overall, PT51IX08, com o nível de confiança de 153.02534; no entanto, como este valor é superior a 100, aplica-se o majorante e o nível de confiança deste agregado passa a 100.

Na Figura 2.1 pode-se ver um dos ficheiros de output do nosso algoritmo com os cálculos realizados anteriormente.

## 2. PRIMEIRA FÓRMULA

42301241	Best Over	PT51IX08	6	4	false	ARR_CARI 72.0	0.0	24.48	114.11625	28.0	10.90909	153.02534	false
42301241	Best Over	PT51IX08	6	4	false	MKW_PLA 93.625	84.0	58.61625	114.11625	28.0	10.90909	153.02534	false
42301241	Best Over	PT51IX08	6	4	false	NLABS_VF 94.0	0.0	31.020000	114.11625	28.0	10.90909	153.02534	false
42301241	Best Front	PT51IX08	6	0	false	ARR_CARI 72.0	0.0	24.48	99.802500	0.0	10.28571	110.08821	false
42301241	Best Front	PT51IX08	6	0	false	MKW_PLA 98.25	36.0	44.3025	99.802500	0.0	10.28571	110.08821	false
42301241	Best Front	PT51IX08	6	0	false	NLABS_VF 94.0	0.0	31.020000	99.802500	0.0	10.28571	110.08821	false
42301241	Best Front	PT51IX08	1	0	false	ARR_CARI 45.0	0.0	15.3	15.3	0.0	0.0	15.3	false
42301241	Best Rear	PT51IX08	0	4	false	MKW_PLA 89.0	36.0	41.25	41.25	0.0	0.0	41.25	false

Figura 2.1: Excerto do Ficheiro do Ficheiro de Output com Cálculos

## 2.2 Primeira Execução

Na primeira execução do algoritmo, para o conjunto das 13921 transações, obtiveram-se 240 erros (equivalente a aproximadamente 1.7% das transações). Utilizaram-se os seguintes valores para as nossas variáveis:

- Photo Matching Factor = 2
- FrontRear Matching Factor = 7
- Engine Matching Factor = 5
- Peso do motor ARH = 0.33
- Peso do motor MKW = 0.33
- Peso do motor NLABS = 0.34
- differentCharsThreshold = 2
- differentCharsMinConfidenceThreshold = 35
- uso das fotos de overview para transações com nível de confiança < 50

Após uma execução do algoritmo, obtém-se um conjunto de outputs. Um destes outputs contém o número de transações erradas e o número total de transações por cada intervalo de níveis de confiança; assim sendo, representa-se esta informação em histogramas. No caso desta execução, o histograma está representado na Figura 2.2

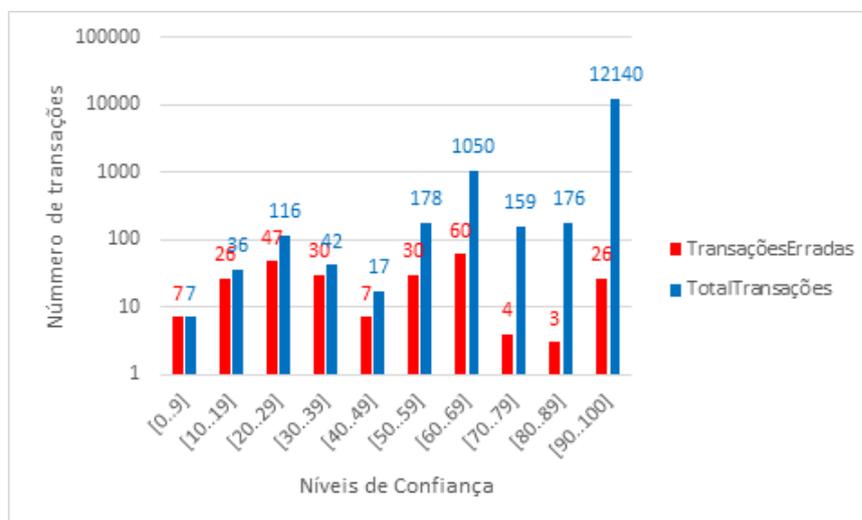


Figura 2.2: Histograma dos Resultados da 1ª Execução do Algoritmo

Inicialmente, como primeira abordagem, geraram-se resultados com um conjunto diverso de parâmetros de modo a entender o efeito dos parâmetros no algoritmo. Isto permitiu diminuir o número de erros para 201, resultado que passou a ser considerado como resultado base. Durante este processo, ao analisar alguns erros, observou-se que o motor da MKW, em certas situações, identificava matrículas sem critério. Este erro levou a uma atualização do motor, e consequentemente atualizaram-se os conhecimentos por parte deste motor guardados no ficheiro de input. Foram também realizadas mais duas

## 2. PRIMEIRA FÓRMULA

atualizações: uma para o motor da NLABS, onde se acrescentou ao motor sintaxes de matrículas de alguns países estrangeiros que circulam no nosso país mais frequentemente, e uma outra atualização que visa disponibilizar a fotografia de overview para todas as transações que fazem parte do nosso conjunto de transações. A ideia era estudar o comportamento de cada uma das atualizações em separado, e no final juntar as três atualizações no mesmo ficheiro de input.

Como já foi referido anteriormente, os vários boosts foram criados para aumentar o nível de confiança conforme a verificação das condições dos mesmos. Um dos factores que intuitivamente tem mais importância é o Front-Rear, pois o facto de a mesma matrícula ser reconhecida tanto em fotos frontais como traseiras deve significar que esta é uma boa candidata a ser escolhida. Decidiu-se então analisar as transações onde as condições deste boost não se verificavam. Verificado o número transações em que Front-Rear Matching Factor tinha um valor nulo, foram encontradas 2332 transações distribuídas pelas diversas classes de veículos, como podemos ver na tabela 2.4. Visto que nesta projecto a antenna não está incluí, para se obter a informação da classe dos veículos das transações em estudo, foi necessário cruzar a informação com outro sistema de modo a obter esta informação. Nesta tabela, classe 0 indica os veículos que não tinham identificador.

Tabela 2.4: Tabela do Número de Transações com Front-Rear Matching Factor Nulo por Classe de Veículo

<b>Nº de Transações</b>	<b>Frequência Relativa (%)</b>	<b>Classe</b>
22	0,94	0
145	6,22	1
138	5,92	2
61	2,62	3
1463	62,74	4
103	4,42	5

É possível ver que, como esperado, a maioria dos casos em que não existem pares de fotografias são de transações associadas a veículos de classe 4.

De seguida pegou-se nas transações das classes 1 e 2 – visto que nestas classes o fator Front-Rear deveria ser não nulo e o número destas transações é relativamente alto – e estudou-se a possível existência de algum padrão de erros a certa altura do dia, não se tendo verificado a existência de qualquer padrão. Por fim, pegou-se em 50 transações destas duas classes e estudou-se o motivo para estas apresentarem factor Front-Rear nulo, tendo-se obtido os resultados apresentados na tabela 2.5.

Tabela 2.5: Tabela com Número de Ocorrência e o Motivo

<b>Nº de Ocorrências</b>	<b>Motivo</b>
22	Veículos de classe 4 mal classificados
6	Existência de fotografias com matrículas escuras
7	Erros na altura certa da captação da fotografia
7	Matrículas ligeiramente obstruída
4	Motas mal classificados
2	Matrícula apresenta manchas
2	Outros casos

É possível observar que a ocorrência mais frequente nesta amostra é existirem veículos de classe 4 mal classificados. Verificou-se também outras ocorrências frequentes neste conjunto de estudo: a

existência de fotografias em que a matrícula se encontra escura, fotografias onde houve um erro na altura da captura da fotografia, e fotografias onde a matrícula se encontra ligeiramente obstruída.

### 2.3 Atualizações do Sistema e Otimização das Variáveis

Como foi visto anteriormente, surgiu a necessidade de fazer algumas atualizações ao nosso sistema e no fim juntá-las. Enquanto as atualizações se realizavam, decidiu-se voltar a realizar uma otimização das nossas variáveis, no entanto desta vez de uma forma não aleatória. O mecanismo utilizado foi, pegando nos valores utilizados no momento, fixaram-se todas as variáveis excepto uma; esta que não se encontrava fixa, reduziu-se e aumentou-se uma unidade e viu-se que alteração diminuía mais o número de erros; caso não houvesse nenhuma diminuição, fixava-se o valor dessa variável e aplicava-se o mesmo método para outra, até termos aplicado este processo aos pesos dos motores e aos matching factors. Este tipo de otimização leva-nos a um mínimo, mas não sabemos se é local ou absoluto.

#### 2.3.1 Atualização no Motor MKW

Tal como foi referido anteriormente, reparou-se que o motor da MKW estaria a ter um comportamento anormal; este comportamento consistia no reconhecimento de matrículas sem qualquer semelhança com a matrícula existente nas fotografias, e também no reconhecimento de matrículas aleatórias em fotografias onde não existiam matrículas, ou as mesmas eram impossíveis de serem lidas. Além disto, notou-se também que, quando o motor não conseguia reconhecer uma matrícula, este retornava “NO-PLATE”, algo que o algoritmo não estava preparado para lidar. Ao aplicar a atualização no motor, pegou-se no ficheiro de input inicial e substituiu-se os reconhecimentos antigos do motor MKW pelos novos. Ao eliminarmos estes dois erros, reduzimos o número de erros para 102 em 13905 transações. O número de transações foi reduzido, pois existiam 16 transações onde o único reconhecimento que existia era “NOPLATE” (que passam a ser consideradas transações erradas).

Aplicando o processo de otimização previamente referido, alcançou-se o valor ótimo, diminuindo o número de erros para 91 em 13905, aproximadamente 0.7%, com a seguinte configuração:

- Photo Matching Factor = 12
- FrontRear Matching Factor = 3
- Engine Matching Factor = 10
- Peso do motor ARH = 0.46
- Peso do motor MKW = 0.34
- Peso do motor NLABS = 0.20
- differentCharsThreshold = 2
- differentCharsMinConfidenceThreshold = 70
- uso das fotos de overview para transações com nível de confiança < 50

A Figura 2.3 representa o histograma dos resultados para a configuração vista anteriormente:

## 2. PRIMEIRA FÓRMULA

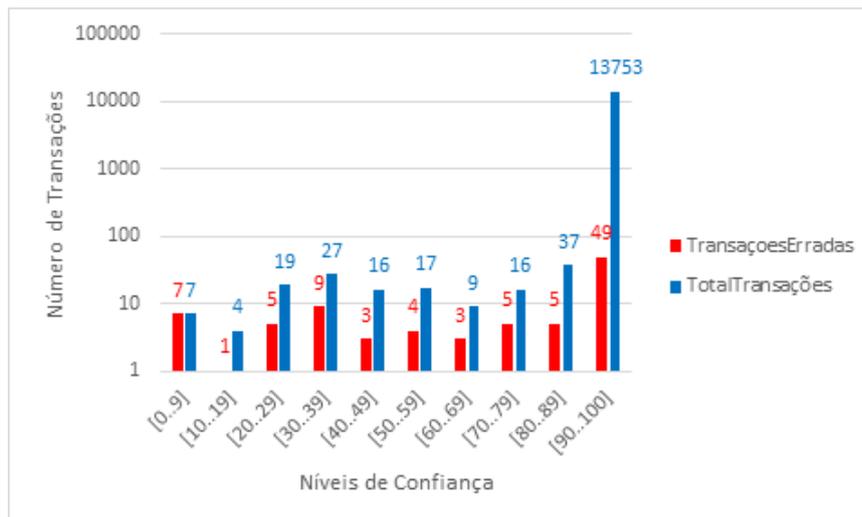


Figura 2.3: Histograma dos Resultados para a Execução dos Valores Otimizados da Atualização do Motor MKW

### 2.3.2 Atualização no Motor NLABS

De modo a aumentar a qualidade dos reconhecimentos em matrículas de veículos estrangeiros, foi realizada uma atualização ao motor NLABS, na qual foram introduzidas novas sintaxes de matrículas referentes a alguns países estrangeiros. O motor já tinha as sintaxes de alguns países; no entanto, foram adicionadas outras, conforme o tráfego de veículos estrangeiros registado no conjunto de dados. Esta atualização foi estudada anulando a atualização relativa ao motor da MKW (isto para compararmos os resultados obtidos com os originais).

Considerando a nova versão do ficheiro de input, aplicando o mesmo método de otimização utilizado na atualização anterior, atingiram-se 127 erros em 13921 transações, aproximadamente 0.9%, utilizando a seguinte configuração:

- Photo Matching Factor = 31
- FrontRear Matching Factor = 28
- Engine Matching Factor = 20
- Peso do motor ARH = 0.46
- Peso do motor MKW = 0.19
- Peso do motor NLABS = 0.36
- differentCharsThreshold = 2
- differentCharsMinConfidenceThreshold = 70
- uso das fotos de overview para transações com nível de confiança < 50

O histograma da Figura 2.4 representa a distribuição dos resultados:

## 2.3 Atualizações do Sistema e Otimização das Variáveis

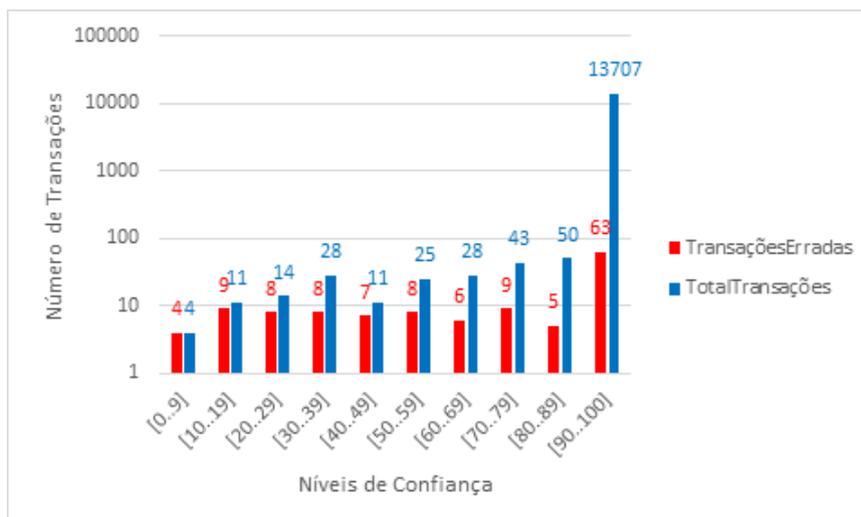


Figura 2.4: Histograma dos Resultados para a Execução dos Valores Otimizados da Atualização do Motor NLABS

### 2.3.3 Atualização das Fotografias de Overview

O input inicial de dados apenas contém algumas fotografias de overview. Esta atualização visa completar o ficheiro de input com todas as fotografias de overview que existirem, para todas as transações em estudo. No entanto apenas existem 13139 transações com fotografias de overview; as restantes, devido a problemas no sistema, não contém fotografias de overview associadas.

Voltando a aplicar o método de otimização utilizado nas atualizações prévias, mas considerando apenas esta atualização, obtivemos 136 erros em 13921 transações, aproximadamente 1%, utilizando a seguinte configuração:

- Photo Matching Factor = 21
- FrontRear Matching Factor = 24
- Engine Matching Factor = 10
- Peso do motor ARH = 0.40
- Peso do motor MKW = 0.30
- Peso do motor NLABS = 0.30
- differentCharsThreshold = 2
- differentCharsMinConfidenceThreshold = 70
- uso das fotos de overview para transações com nível de confiança < 50

O histograma da Figura 2.5 representa a distribuição dos resultados:

## 2. PRIMEIRA FÓRMULA

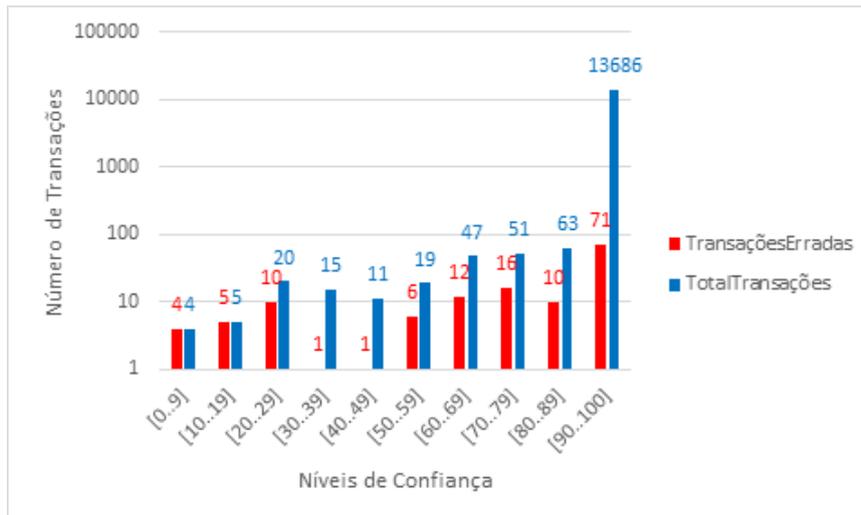


Figura 2.5: Histograma dos Resultados para a Execução dos Valores Otimizados da Atualização das Fotografias de Overview

### 2.3.4 Junção das Atualizações

Realizaram-se três atualizações aos nossos dados e em todas o número de erros diminuiu. A ideia é unir as três atualizações de modo a que esta combinação diminua o maior número de erros possível, e de preferência que este valor seja menor que o das restantes atualizações. Não estava prevista mais nenhuma atualização; então utilizando os dados referentes a esta atualização, serão realizados um conjunto de estudos que não ocorreram nas atualizações anteriores.

Repetimos o processo de otimização, obtendo 93 erros em 13916 transações, aproximadamente 0.7%, utilizando os seguintes valores:

- Photo Matching Factor = 27
- FrontRear Matching Factor = 1
- Engine Matching Factor = 33
- Peso do motor ARH = 0.42
- Peso do motor MKW = 0.29
- Peso do motor NLABS = 0.29
- differentCharsThreshold = 2
- differentCharsMinConfidenceThreshold = 70
- uso das fotos de overview para transações com nível de confiança < 50

O histograma da Figura 2.6 representa a distribuição dos resultados:

## 2.4 Análise de Matching Levels de Erros

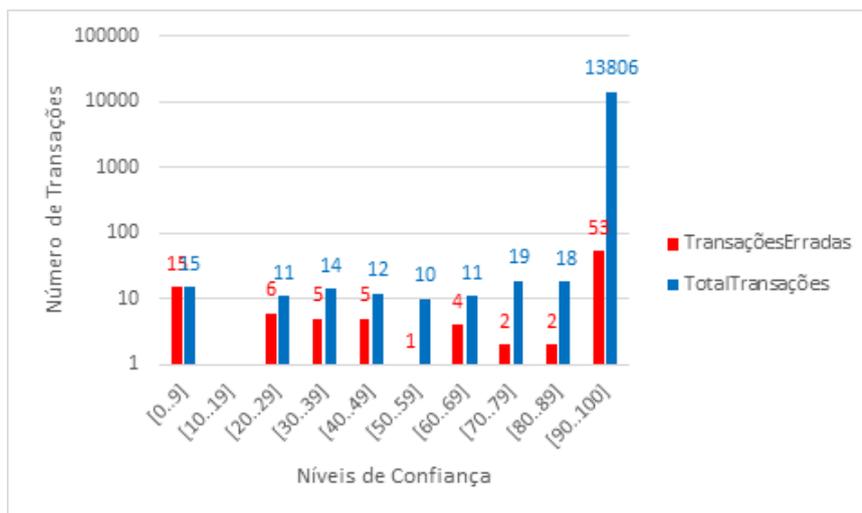


Figura 2.6: Histograma dos Resultados para a Execução dos Valores Otimizados da Junção das Atualizações

Como se pode ver, o número de erros desta junção foi 93, igualando o valor obtido na atualização do motor da MKW. No entanto, como neste caso o número de transações em estudo foi superior, podemos dizer que esta junção apresenta melhores resultados.

## 2.4 Análise de Matching Levels de Erros

De modo a avaliar o comportamento dos matching levels (parcelas da fórmula nas quais figura o matching factor respetivo), aumentou-se o valor de cada um dos parâmetros de modo a que seja garantido que, na presença de pelo menos um matching factor, os níveis de confiança alcancem o valor 100. Analisaram-se os erros, considerando cada uma das atualizações, aumentou-se o valor de cada matching factor para 1000, e realizou-se então uma contagem de quantos casos tinham algum, todos, ou nenhum matching level nulo. Os resultados encontram-se representados na Figura 2.6

Tabela 2.6: Tabela com Número de Erros de Matching Level Nulos por Atualização

Matching Level Nulo	Atualização			
	Motor da MKW	Motor da NLABS	Fotografias de Overview	Junção
Photo	7	22	21	7
Engine	25	50	34	30
Front-Rear	60	90	78	47
Todos	7	21	16	6
Nenhum	15	22	65	15
<b>Total de transações</b>	<b>85</b>	<b>123</b>	<b>153</b>	<b>74</b>

Ora, esta análise mostra-nos que o algoritmo está a ter o comportamento esperado em todas as atualizações. O boost que é nulo mais vezes nos erros é sempre o Front-Rear, o que mostra que a ausência deste boost é a que aparenta ter maior impacto negativo na escolha da matrícula correta. Olhando para os valores da junção das atualizações, podemos ver que os números são quase todos inferiores aos das restantes atualizações, excepto o número referente ao Photo Matching Level nulo e a nenhum matching level nulo (pois apresenta números iguais à atualização do motor da MKW). Portanto, podemos concluir

## 2. PRIMEIRA FÓRMULA

que a atualização que faz a junção aparenta ser a melhor atualização feita, seguida pela atualização ao motor da MKW.

### 2.5 Análise de Erros

Como a esta altura não estava prevista mais nenhuma alteração no input, então pôde-se realizar um estudo dos dados mais aprofundado. Como tal, começou-se por estudar ao pormenor os 93 erros:

- 24 transações de veículos estrangeiros
- 14 transações sem qualquer reconhecimento
- 38 transações com veículos de classe 4
- 11 transações com motas
- 33 transações onde a foto frontal tem problemas de luminosidade
- 20 transações onde o Manual Review está errado
- 40 transações onde a matrícula correta foi identificada, mas não foi a escolhida
- 17 transações onde a matrícula correta não foi reconhecida

Esta análise detalhada dos erros revela que existem várias manual reviews erradas por erro humano; revela também que existe um erro na implementação do algoritmo, pois os reconhecimentos de fotografias de overview não eram utilizados quando estes reconhecimentos eram os únicos na transação, e também que existe um grande número de erros associados à escolha da matrícula traseira de veículos de classe 4..

Olhando para a Figura 2.6, podemos ver que os erros se encontram distribuídos por todos os intervalos de níveis de confiança, algo que não é desejável pois, tal como foi referido na introdução, no caso da existência de erros é desejável que estes se encontrem abaixo de um certo nível de confiança.

### 2.6 Estudo dos Thresholds

Até este ponto, não foi feito nenhum estudo sobre o a alteração dos thresholds (`differentCharsThreshold` e `differentCharsMinConfidenceThreshold`). No entanto, a variação da confiança mínima frontal pode eliminar alguns erros. Primeiramente, olhando para as 13902 transações, contam-se quantas acertam, ou não, caso seja escolhido o melhor overall, ou quantas acertam, ou não, caso seja escolhido a melhor frontal, como podemos ver na tabela 2.7.

Tabela 2.7: Tabela com o Número de Melhor Overall/Frontal Certo/Errado

<b>Melhor Overall Certo</b>	13377
<b>Melhor Overall Errado</b>	525
<b>Melhor Frontal Certa</b>	13584
<b>Melhor Frontal Errada</b>	318

Como se pode ver, existem mais transações com a melhor frontal certa que transações com o melhor overall. Isto indica que existe uma possível otimização dos thresholds para que se consiga conjugar

## 2.7 Introdução da Regra de Sintaxe – Simulação Vs Algoritmo

a escolha entre o melhor overall e a melhor frontal. Analisando os níveis de confiança destes mesmo agregados vai-se simular a procura da combinação dos thresholds, de modo a que o número de transações que falham na escolha do melhor overall e acertam na escolha frontal seja maior, e que o número de transações que acertam a escolha do melhor overall e falham na escolha do melhor frontal seja menor. Obtiveram-se os resultados indicados na tabela 2.8:

Tabela 2.8: Tabela de Variação dos Thresholds e Número de erros

<b>differentCharsThreshold</b>	1	2
<b>differentCharsMinConfidenceThreshold</b>	76	67
<b>Total de erros</b>	61	49

Olhando para os valores da tabela 2.8, pode-se concluir que nestas 13902 transações, ao testar as condições para escolha de melhor matrícula frontal, considerando `differentCharsThreshold=2` e `differentCharsMinConfidenceThreshold = 67`, atinge-se um valor mínimo de erros.

## 2.7 Introdução da Regra de Sintaxe – Simulação Vs Algoritmo

Como foi referido na introdução, os veículos de classe 4 apresentam a matrícula do veículo na matrícula frontal, enquanto que a matrícula traseira apresenta a referente ao reboque e, como sabemos, esta apresenta uma sintaxe própria. Como tal, decidimos simular a aplicação de uma regra de sintaxe, regra essa que ao identificar a sintaxe de uma matrícula de um reboque, sinaliza ao algoritmo que deverá escolher a matrícula frontal. No entanto, o nosso algoritmo não considera as matrículas da forma XX-0000, visto que este conjunto de matrículas pode representar tanto reboques como veículos. Realizámos três simulações da aplicação desta regra. A primeira utilizando apenas a regra de sintaxe como única condição para a escolha da matrícula frontal como matrícula escolhida para a transação em estudo, tendo-se obtido 86 erros em 13902 transações. Nas restantes duas simulações incluímos também a utilização dos thresholds, uma com as variáveis `differentCharsThreshold=1` e `differentCharsMinConfidenceThreshold=76` que resultou em 56 erros, e na outra simulação considerámos as variáveis `differentCharsThreshold=2` e `differentCharsMinConfidenceThreshold=67`, obtendo 48 erros; ambas as simulações tiveram em estudo 13902 transações. Estas simulações não têm em consideração a utilização das fotografias de overview.

Correndo o algoritmo, incluindo a utilização da regra de sintaxe e utilizando as variações ótimas das variáveis `differentCharsThreshold` e `differentCharsMinConfidenceThreshold`, obtivemos os resultados representados na tabela 2.9:

Tabela 2.9: Tabela de Variação dos Thresholds e Número de erros Utilizando Regra de Sintaxe

<b>differentCharsThreshold</b>	1	2
<b>differentCharsMinConfidenceThreshold</b>	76	67
<b>Total de erros</b>	57	58

Como o algoritmo já tem em consideração a utilização das fotografias de overview, o número de erros dos resultados do algoritmo apresenta-se diferente dos valores obtidos pela simulação. No final do projeto, ao analisar o código do algoritmo, veio-se a descobrir que a escolha da matrícula frontal só acontece quando: a diferença de caracteres entre as melhores matrículas é maior que o `differentCharsThreshold`, e o nível de confiança da melhor matrícula frontal é maior que `differentCharsMinConfidenceThreshold`.

## 2. PRIMEIRA FÓRMULA

Nesta simulação pensava-se que quando a diferença de caracteres era igual ao `differentCharsThreshold` e o nível de confiança da matrícula era igual ao `differentCharsMinConfidenceThreshold`, então era escolhida a matrícula frontal. Assim sendo, este pode ser mais um motivo para os números de erros dos resultados da simulação serem diferentes dos resultados do algoritmo.

Nas figuras 2.7 e 2.8 podemos ver os diagramas dos resultados referidos na tabela 2.9.

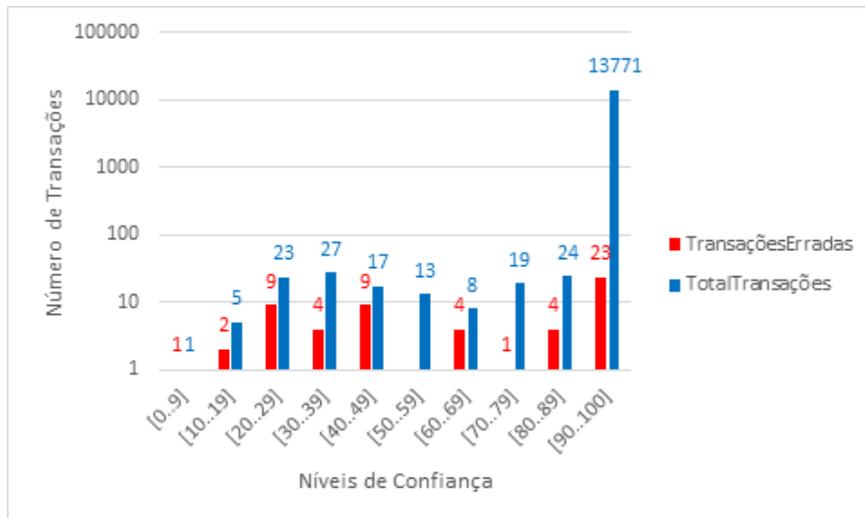


Figura 2.7: Histograma dos Resultados com `differentCharsThreshold=1` e `differentCharsMinConfidenceThreshold=76`

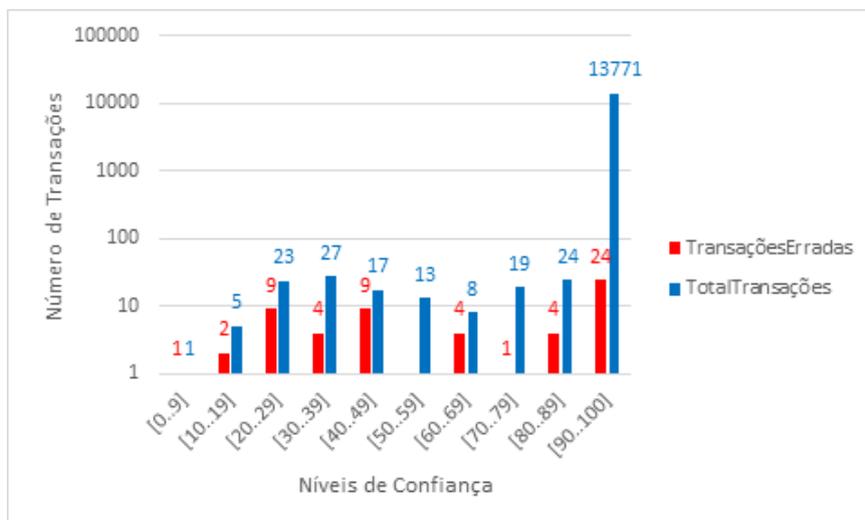


Figura 2.8: Histograma dos Resultados com `differentCharsThreshold=2` e `differentCharsMinConfidenceThreshold=67`

Olhando para estes últimos resultados obtidos, em 13908 transações, o menor número de erros obtido foi 57, equivalente a 0.4%, dispersos por todos os níveis de confiança. Este foi o menor número de erros obtido utilizando o algoritmo e é um valor satisfatório. No entanto, temos 13771 transações no intervalo de confiança [90..100], equivalente a 99%, pelo que se pode concluir que estes resultados apresentam uma dispersão bastante reduzida. Quanto às transações com nível de confiança 100, não se sabe se estas têm realmente nível de confiança com esse valor ou se sofreram a majoração por parte do algoritmo.

## Capítulo 3

# Nova Fórmula

De modo a eliminar os defeitos que vimos anteriormente, e a aproveitar ao máximo a informação que dispomos tanto de input, que já era utilizada, como de output do algoritmo anterior, realizou-se uma reformulação do nosso algoritmo. Reformulação essa que tem como base a substituição dos pesos dos motores e dos matching factors, que previamente eram constantes, por funções que vão receber o peso dos motores ou condições referentes aos boosts e vão retornar um nível de confiança. Estas funções representam a forma como cada factor e motor se relaciona com as matrículas corretas.

### 3.1 Reformulação

Para calcular as funções de cada motor, vamos analisar quantos reconhecimentos o motor acertou para cada nível de confiança, e essa percentagem de sucessos vai passar a ser o nível de confiança que a nossa função nos vai retornar. Por exemplo: se um motor, em 100 reconhecimentos com nível de confiança de 50, acertar apenas 30, então temos 30% de sucesso; logo, a função quando recebe o nível de confiança de 50 para esse motor, vai atualizá-lo para 30.

A função referente ao Photo Matching Boost, vai retornar um nível de confiança conforme o número de fotografias que identificaram a matrícula do agregado em causa, número esse que varia entre 1 a 8 (ignoramos a fotografia de overview). Olhando para todos os agregados e vendo para cada número possível de fotografias, vamos calcular a percentagem de sucesso que vai ser o nível de confiança retornado pela função.

Quanto à função referente ao Engine Matching Boost, esta vai retornar um nível de confiança conforme o número de motores que identificaram a matrícula do agregado em causa, número este que varia de 1 a 3. Olhando para todos os agregados e vendo o número de motores, calcula-se a percentagem de sucesso para cada número de motores que, tal como nas outras funções, será o nível de confiança retornado.

Referente ao Front-Rear Matching Boost, a nossa função vai retornar um nível de confiança conforme o número de pares de fotografias frontais e traseiras que identificam a matrícula do agregado em causa, o número de pares varia entre 0 e 4. Repetimos o processo utilizado nas 2 funções anteriormente referidas, adaptando ao número de pares de fotografias. Tendo estas funções, falta saber qual a melhor maneira de as relacionar, para um melhor cálculo de um nível de confiança final e consequentemente um menor número de transações erradas.

A ideia inicial consistia em atualizarmos *a priori* os níveis de confiança dos reconhecimentos, antes da agregação, de modo a que na criação dos agregados seja calculado o nível médio de confiança do agregado já com os níveis de confiança dos reconhecimentos atualizados. Tendo em conta o número

### 3. NOVA FÓRMULA

de fotografias, o número de motores, e o número de pares de fotografias frontais e traseiras, para o agregado em estudo, as funções referentes a estas condições retornam (cada uma) um nível de confiança. O agregado tem agora quatro níveis de confiança associados e, como primeira abordagem, o nível de confiança final é obtido através de uma média destes quatro valores.

## 3.2 Funções de Confiança

### 3.2.1 Funções de Confiança dos Motores

As funções de atualização dos níveis de confiança dos reconhecimentos dos motores são constituídas por um conjunto de pontos; então aplicaram-se regressões lineares de modo a obter-se expressões que representem o mais corretamente possível o comportamento dos dados. Considera-se a variável  $nConf$  como o nível de confiança que a função retorna, e  $oConf$  o nível de confiança que a função recebe e atualiza.

#### 3.2.1.1 ARH

Para a função do motor da ARH obteve-se o gráfico da Figura 3.1 e a função 3.1:

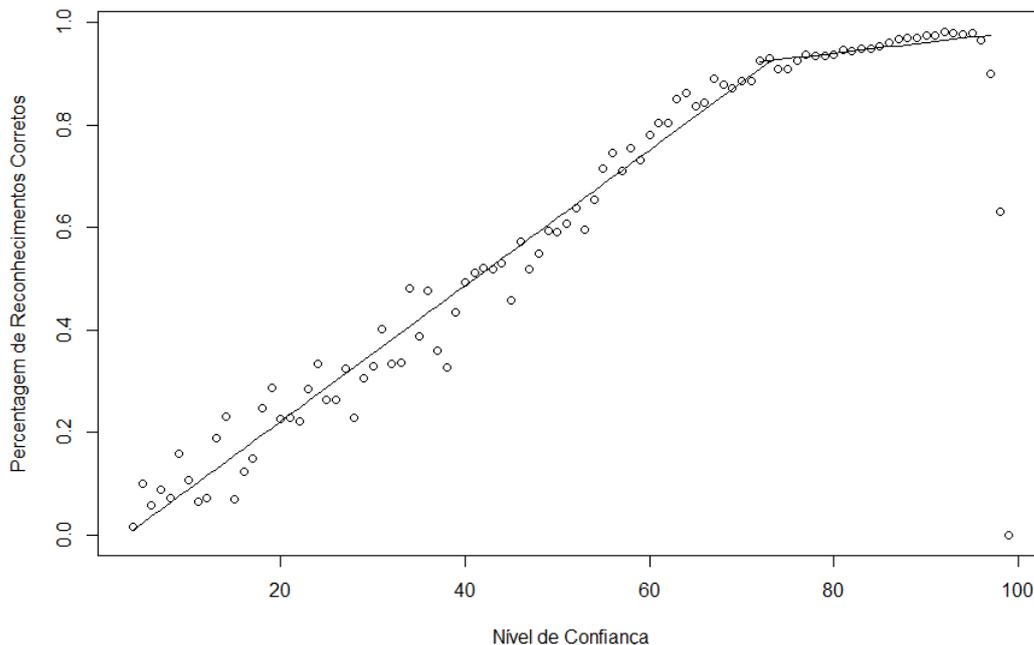


Figura 3.1: Gráfico da Função de Reconhecimentos do motor ARH

$$nConf = \begin{cases} -4.3 - 1.3 * oConf, & \text{se } oConf \leq 73 \\ 77.3 + 0.2 * oConf, & \text{se } oConf > 73 \end{cases} \quad (3.1)$$

#### 3.2.1.2 MKW

Para a função do motor da MKW obteve-se o gráfico da Figura 3.2 e a função 3.2:

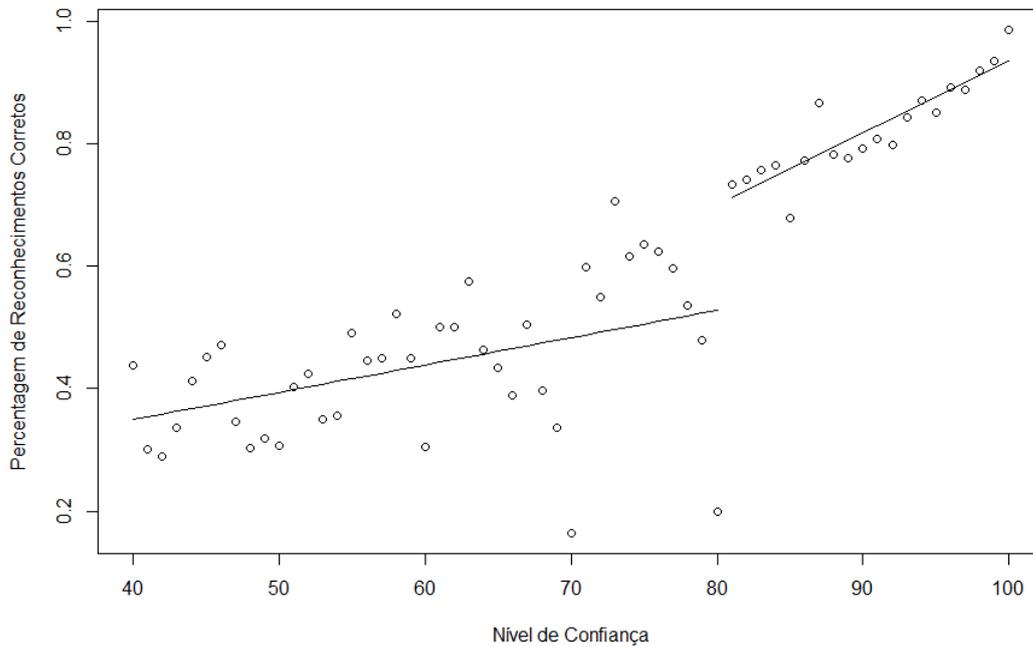


Figura 3.2: Gráfico da Função de Reconhecimentos do motor MKW

$$nConf = \begin{cases} 17.1 + 0.4 * oConf, & \text{se } oConf \leq 80 \\ -28 + 1.2 * oConf, & \text{se } oConf > 80 \end{cases} \quad (3.2)$$

No conjunto de pontos com níveis de confiança abaixo de 80, a nuvem de pontos apresenta uma maior dispersão, no entanto a regressão linear consegue representar estes pontos de uma forma significativa.

#### 3.2.1.3 NLABS

Para a função do motor da NLABS obteve-se o gráfico da Figura 3.3 e função 3.3.:

### 3. NOVA FÓRMULA

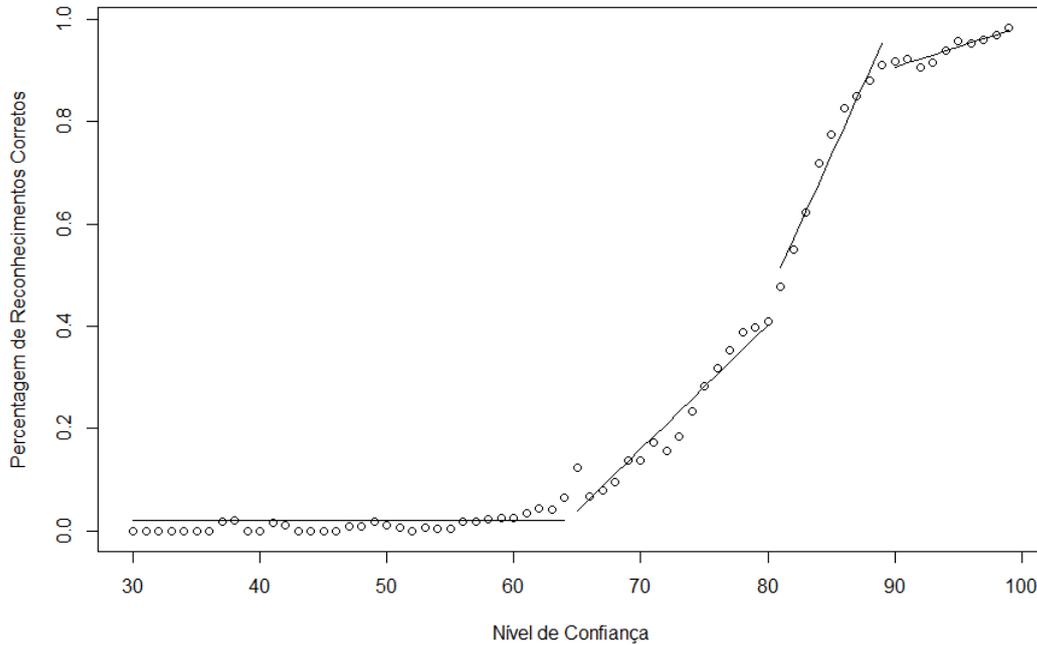


Figura 3.3: Gráfico da Função de Reconhecimentos do motor NLABS

$$nConf = \begin{cases} 2, & \text{se } oConf < 65 \\ -153.9 + 2.4 * oConf, & \text{se } 65 \leq oConf \leq 80 \\ -392.8 + 5.5 * oConf, & \text{se } 81 \leq oConf \leq 89 \\ 18.8 + 0.8 * oConf, & \text{se } oConf > 89 \end{cases} \quad (3.3)$$

#### 3.2.2 Funções de Confiança dos Matching Factors

As funções referentes aos boosts, visto que apenas podem tomar um conjunto limitado de valores, vão ser constituídas por ramos. Neste primeiro estudo não estão incluídas as fotografias de overview pois a maioria das transações não utiliza as mesmas.

##### 3.2.2.1 Photo Matching Factor

Considerando  $N$  o número de fotografias que identificam o agregado e  $pConf$  o nível de confiança retornado pela função, para a Função de Photo Matching Factor os resultados estão representados no gráfico da Figura 3.4, convertendo-se na função 3.4:

### 3.2 Funções de Confiança

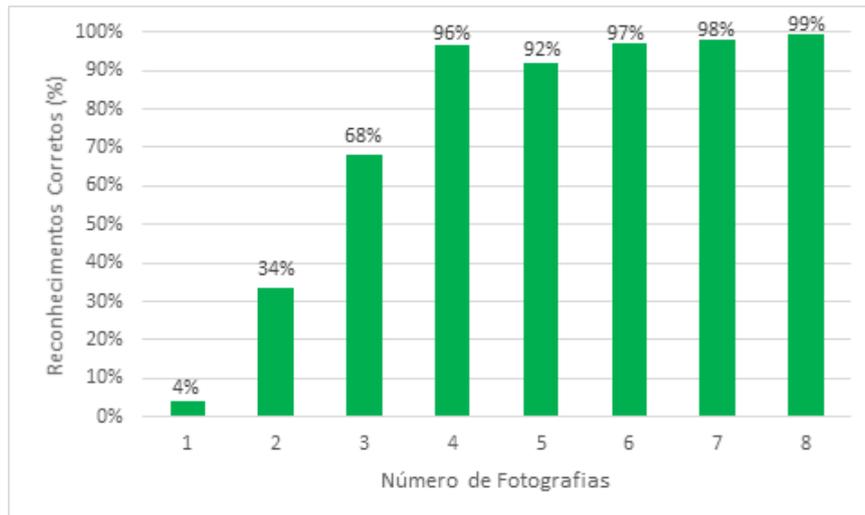


Figura 3.4: Gráfico da Função de Photo Matching Factor

$$pConf = \begin{cases} 4, & se \ N = 1 \\ 34, & se \ N = 2 \\ 68, & se \ N = 3 \\ 96, & se \ N = 4 \\ 92, & se \ N = 5 \\ 97, & se \ N = 6 \\ 98, & se \ N = 7 \\ 99, & se \ N = 8 \end{cases} \quad (3.4)$$

#### 3.2.2.2 Engine Matching Factor

Considerando  $N$  o número de motores que identificam o agregado e  $eConf$  o nível de confiança retornado pela função, para a Função de Engine Matching Factor os resultados estão representados no gráfico da Figura 3.5, convertendo-se na função 3.5:

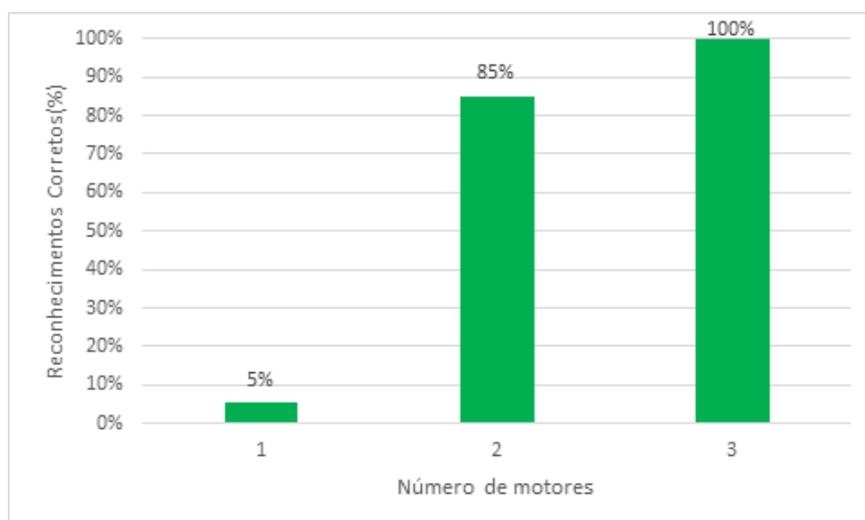


Figura 3.5: Gráfico da Função de Engine Matching Factor

### 3. NOVA FÓRMULA

$$eConf = \begin{cases} 5, & \text{se } N = 1 \\ 85, & \text{se } N = 2 \\ 99, & \text{se } N = 3 \end{cases} \quad (3.5)$$

#### 3.2.2.3 Front-Rear Matching Factor

Considerando  $N$  o número de pares de fotografias frontais e traseiras que identificam o agregado e  $frConf$  o nível de confiança retornado pela função, para a Função de Front-Rear Matching Factor os resultados estão representados no gráfico da Figura 3.6, convertendo-se na função 3.6:

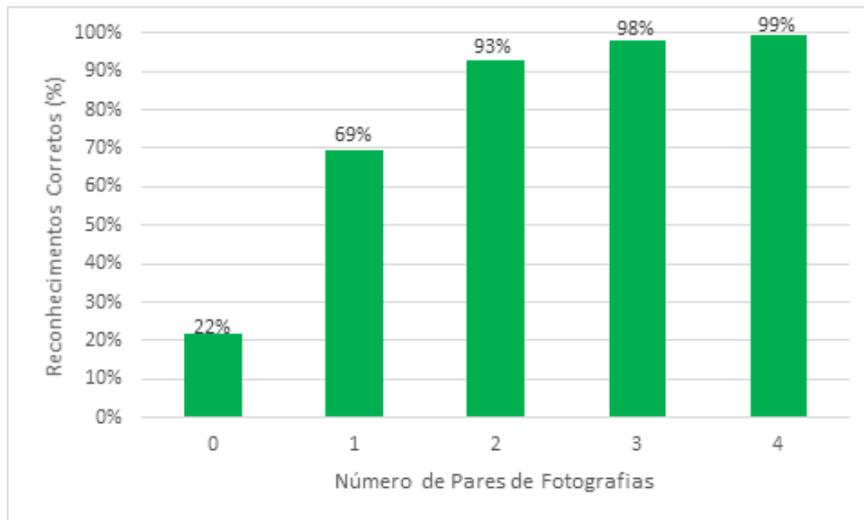


Figura 3.6: Gráfico da Função de Front-Rear Matching Factor

$$frConf = \begin{cases} 22, & \text{se } N = 0 \\ 69, & \text{se } N = 1 \\ 93, & \text{se } N = 2 \\ 98, & \text{se } N = 3 \\ 99, & \text{se } N = 4 \end{cases} \quad (3.6)$$

### 3.3 Primeira Execução

Correndo o algoritmo utilizando a nova fórmula, considerando que o nível de confiança final de um agregado é calculado através da média aritmética, que  $differentCharsThreshold=2$  e  $differentCharsMinConfidenceThreshold=67$ , e que a utilização das fotografias de overview ocorre para transações com níveis de confiança menores que 50, obtiveram-se 68 erros em 13907 transações, o que corresponde aproximadamente a uma taxa de erro de 0.5%.

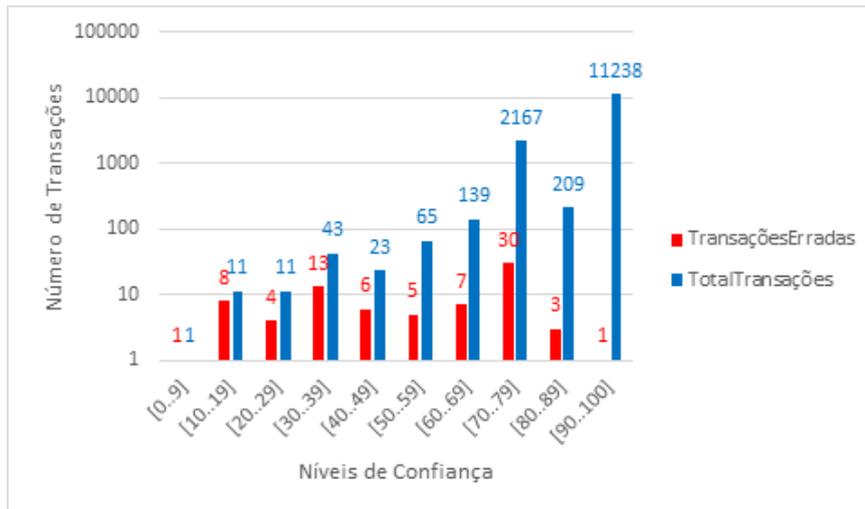


Figura 3.7: Histograma com os Resultados da Primeira Execução c/ nova Fórmula

### 3.4 Estudo dos Thresholds

Como vimos no estudo da primeira fórmula, a alteração das variáveis `differentCharsThreshold` e `differentCharsMinConfidenceThreshold` pode resultar numa diminuição dos erros, e como tal esta questão foi a primeira a ser estudada. Considerando as 13897 que estão no nosso ficheiro de cálculos, vamos ver quantas transações estarão corretas considerando a alteração em causa. Primeiramente viu-se que existem 13067 casos onde tanto a escolha overall como a frontal se encontram corretas e, para cada caso, viu-se também quantos casos eram beneficiados ou prejudicadas por cada variação das nossas variáveis. Variando o `differentCharsThreshold` entre 1 e 3, variou-se o `differentCharsMinConfidenceThreshold` de 0 a 100; o melhor valor obtido foi 13561, atribuindo a `differentCharsThreshold` o valor 2 e a `differentCharsMinConfidenceThreshold` o valor 36. Utilizando estas alterações, ao executar o algoritmo, obteve-se 68 erros em 13907 transações.

### 3.5 Alternativas

Já com a otimização das nossas variáveis feita, estuda-se a necessidade da criação de funções para os Boosts que incluam as fotografias de overview e estuda-se também se será mais adequado calcular o nível de confiança final utilizando uma média aritmética ou uma média ponderada. Considerando `wConf` o nível médio de confiança do agregado, a média ponderada é representada em 3.7:

$$\frac{wConf^2 + pConf^2 + eConf^2 + frConf^2}{wConf + pConf + eConf + frConf} \tag{3.7}$$

Correndo o algoritmo para cada combinação destes casos, a seguinte tabela representa o número de erros obtidos em 13907 transações:

### 3. NOVA FÓRMULA

Tabela 3.1: Tabela com o Número de Erros para cada Variação

	Sem Funções de Overview	Com Funções de Overview
Média Aritmética	68	70
Média Ponderada	127	127

Como se pode ver, o menor número de erros é obtido quando não se utilizam funções próprias para as fotografias de overview e o nível de confiança final de um agregado é calculado através da média aritmética, que é a opção com menos complexidade computacional. No entanto, isto não implica que este seja o método mais correto de relacionar os quatro valores; há a possibilidade de existir um método que obtenha melhores resultados.

### 3.6 Análise de Erros

Na execução anterior, obtivemos 68 erros e realizou-se uma análise aos mesmo, de modo a descobrir os motivos destas ocorrências. Obtivemos a seguinte lista:

- 19 transações são referentes a veículos de classe 4
- 17 transações onde a matrícula correta não é identificada em nenhum agregado
- 4 transações onde a escolha da matrícula frontal induziu em erro
- 3 transações onde apenas o país está errado
- 35 transações onde a matrícula correta não foi escolhida devido ao seu nível de confiança ser demasiado baixo
- 5 transações onde a matrícula escolhida foi a única reconhecida
- 20 transações onde a matrícula correta é estrangeira
- 14 transações onde temos veículos da GNR
- 3 transações com manual review errado
- 2 transações onde a foto de overview utilizada é referente a outro veículo

Com esta análise, ao corrigirem-se as manual reviews reduziram-se 2 erros, e reparou-se que o motor da MKW reconhecia os veículos da GNR como alemães, reconhecimentos esses que foram retirados do nosso input, reduzindo mais 3 erros. Quanto aos veículos de matrícula estrangeira existem duas situações: a primeira consiste no facto de existirem matrículas italianas e francesas com sintaxes semelhantes (estes dois países têm as sintaxes incluídas nos nossos motores), o que resultou em 2 dos nossos erros; a outra situação, são países cujas sintaxes não estão incluídas nos nossos motores, logo estes nunca irão acertar estes reconhecimentos. No caso das 2 transações onde a fotografia de overview não pertence às mesmas, basta-nos eliminar as fotografias do nosso input; uma das transações desaparece, pois apenas continha a fotografia de overview. Quanto ao resto dos erros não há muito que se possa fazer, pois tal como foi visto nos erros da primeira fórmula, existem transações onde as fotografias não permitem identificar as matrículas corretamente.

### 3.6.1 Ajuste de Erros

Alguns destes erros podem ter níveis de confiança elevados (os erros com níveis de confiança acima de 80, eram veículos estrangeiros). De modo a controlar este problema, surge a ideia de aplicar uma modificação dos níveis de confiança de veículos que o algoritmo identifique como estrangeiros, deixando assim de existir erros acima do nível de confiança 80. Esta modificação faz-se multiplicando o valor do nível de confiança de matrículas estrangeiras por 0.8. Pela primeira vez, existe um nível de confiança máximo para os nossos erros, algo que é desejado e pode permitir que, num processo de automatização do algoritmo, apenas transações com níveis de confiança abaixo de 80 sejam sujeitas a um manual review.

Quanto aos erros provenientes de transações de veículos de classe 4, pode-se tentar atualizar a regra de sintaxe. Neste algoritmo utiliza-se a regra de sintaxe para veículos de classe 4 criada na primeira fórmula; no entanto, esta não inclui a sintaxe XX-0000, pelo que se decidiu introduzi-la para ver se esta regra iria prejudicar reconhecimentos de veículos de outras classes com a mesma sintaxe.

## 3.7 Resultados da análise de erros

Aplicando as correções dos erros vistas anteriormente, e a transação das matrículas estrangeiras, ao correr o algoritmo obtiveram-se 60 erros em 13906 transações e o histograma da Figura 3.8:

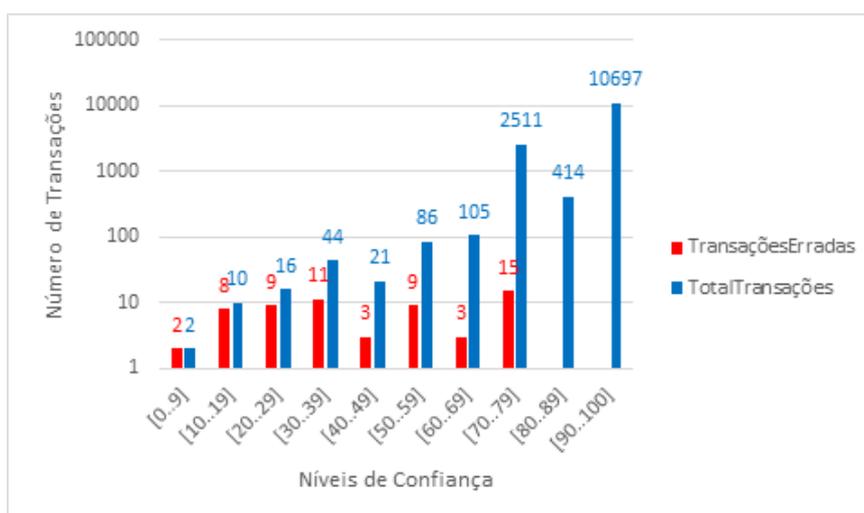


Figura 3.8: Histograma com os Resultados da Execução com Translação de Matrículas Estrangeiras

Ao atualizar-se a regra de sintaxe dos veículos de classe 4, o algoritmo obteve apenas 58 erros em 13906 transações, representados na Figura 3.9:

### 3. NOVA FÓRMULA

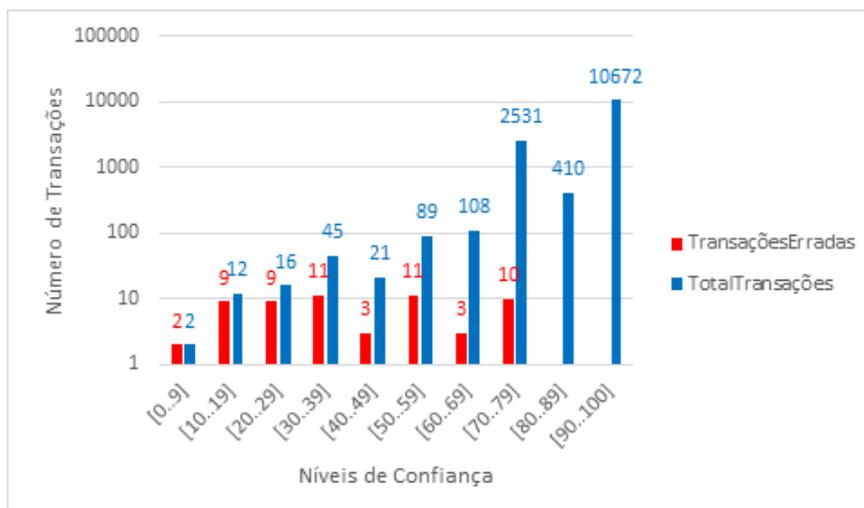


Figura 3.9: Histograma com os Resultados da Execução com Translação de Matrículas Estrangeiras e Regra de Sintaxe Atualizada

## Capítulo 4

# Conclusão

### 4.1 Conclusão

Relativamente aos resultados utilizando a primeira fórmula para o cálculo dos níveis de confiança finais, após a junção das várias atualizações aplicadas sobre os nossos dados de input e a introdução de uma regra de identificação de sintaxe de reboques de veículos de classe 4, o menor número de erros atingido foi 57 em 13902 transações o que equivale a uma taxa de erro de 0.4%, enquanto que no início deste estágio tínhamos 240 erros em 13921 transações equivalendo a uma taxa de erro de 1.7%.

Os erros encontram-se dispersos ao longo de todos os intervalos de níveis de confiança, o que impossibilita a criação de um nível de confiança máximo para as transações passarem por um processo de manual review. Devido à falta de limitações nos valores que os matching factors podem tomar, durante a execução do algoritmo, os níveis de confiança podem atingir valores superiores a 100; quando o algoritmo seleciona uma matrícula em que esta situação aconteça, aplica o majorante 100 e existe uma perda de informação – tal como os dados nos transmitem, temos 13771 resultados, que representa 99% do total, compreendidos no intervalo de confiança [90,100]. Além disso, esta fórmula encontra-se otimizada para este conjunto de dados, possivelmente para um outro conjunto de dados teria de ser feita outra otimização parâmetro a parâmetro.

De modo a evitar estas situações, surge uma reformulação da fórmula do cálculo dos níveis de confiança finais que visa majorar o cálculo dos níveis de confiança no valor 100, proporcionar uma maior dispersão dos nossos dados e a utilização de funções que, conforme o comportamento dos dados, ou atualizam os níveis de confiança dos reconhecimentos de cada motor, ou retornam um nível de confiança conforme o comportamento de cada matching level. Utilizando a média aritmética dos resultados obtidos através das funções, obtivemos 58 erros em 13906 transações, o equivalente a uma taxa de erros de 0.4%. Neste caso, apenas temos erros abaixo do nível de confiança 80, e só as transações abaixo deste valor necessitarão de um manual review. Esta versão do algoritmo é mais fácil de ser adaptado a outro conjunto de dados, pois basta atualizar as funções, tornando-se também melhor no âmbito da automatização do sistema.

### 4.2 Desenvolvimento do estudo

Ao longo do projeto realizaram-se várias correções em erros tanto nos ficheiros de input como nos algoritmos. No entanto, não se voltou atrás para se obter resultados anteriores atualizados. Assim sendo, recomenda-se que se volte a correr o algoritmo, pelo menos com a ultima versão da primeira fórmula, mas utilizando o ficheiro de input mais recente.

#### 4. CONCLUSÃO

No âmbito do desenvolvimento do projeto, devem procurar-se padrões de comportamento anormais a nível das câmaras e realizar estudos onde devem ser utilizadas combinações de diferentes variáveis; por exemplo, o comportamento das câmaras ao longo das horas do dia. Conforme os resultados, pode-se adaptar as funções de confiança às câmaras que o justifiquem.

É também possível que a criação de novos boosts possa trazer resultados benéficos para o algoritmo. Para se relacionar as funções, utilizámos uma média simples mas não existem garantias que esta é a melhor maneira. Talvez se deva estudar fórmulas condicionais de modo a privilegiar alguns factores em detrimento de outros. Outra possível abordagem seria alterar o conjunto de transações em estudo e ver o comportamento do algoritmo.

Quando se atingirem resultados consistentes para diferentes tipos de dados, deve-se proceder à automatização da atualização das funções.